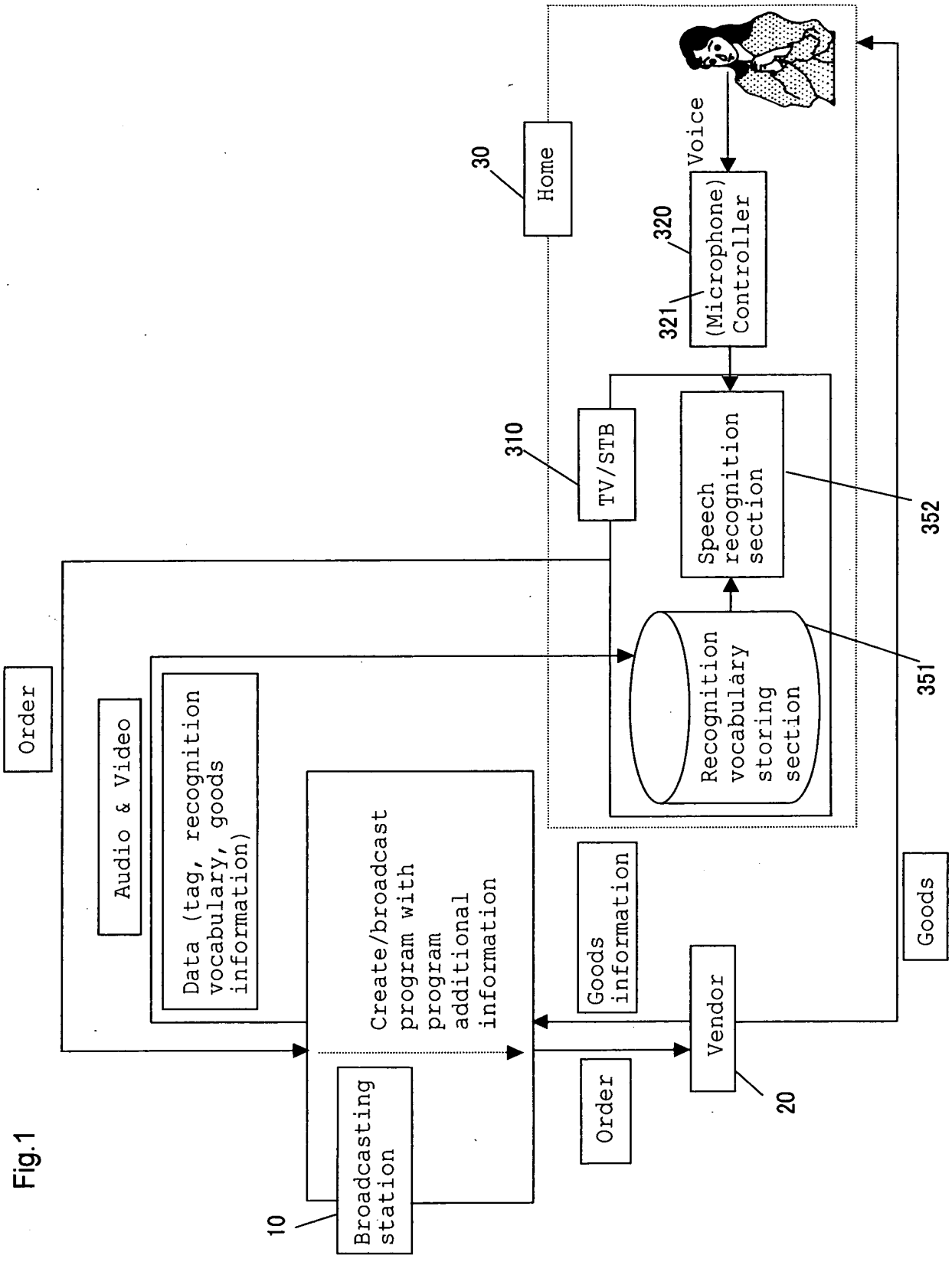

Order
Audio & Video
Data (tag, recognition vocabulary, goods information)
Create/broadcast program with program additional information
Broadcasting station
10
Order
Goods information
Vendor
20
Goods
Home
30
Voice
320
(Microphone) Controller
321
310
TV/STB
Speech recognition section
352
Recognition vocabulary storing section
351

Fig.1

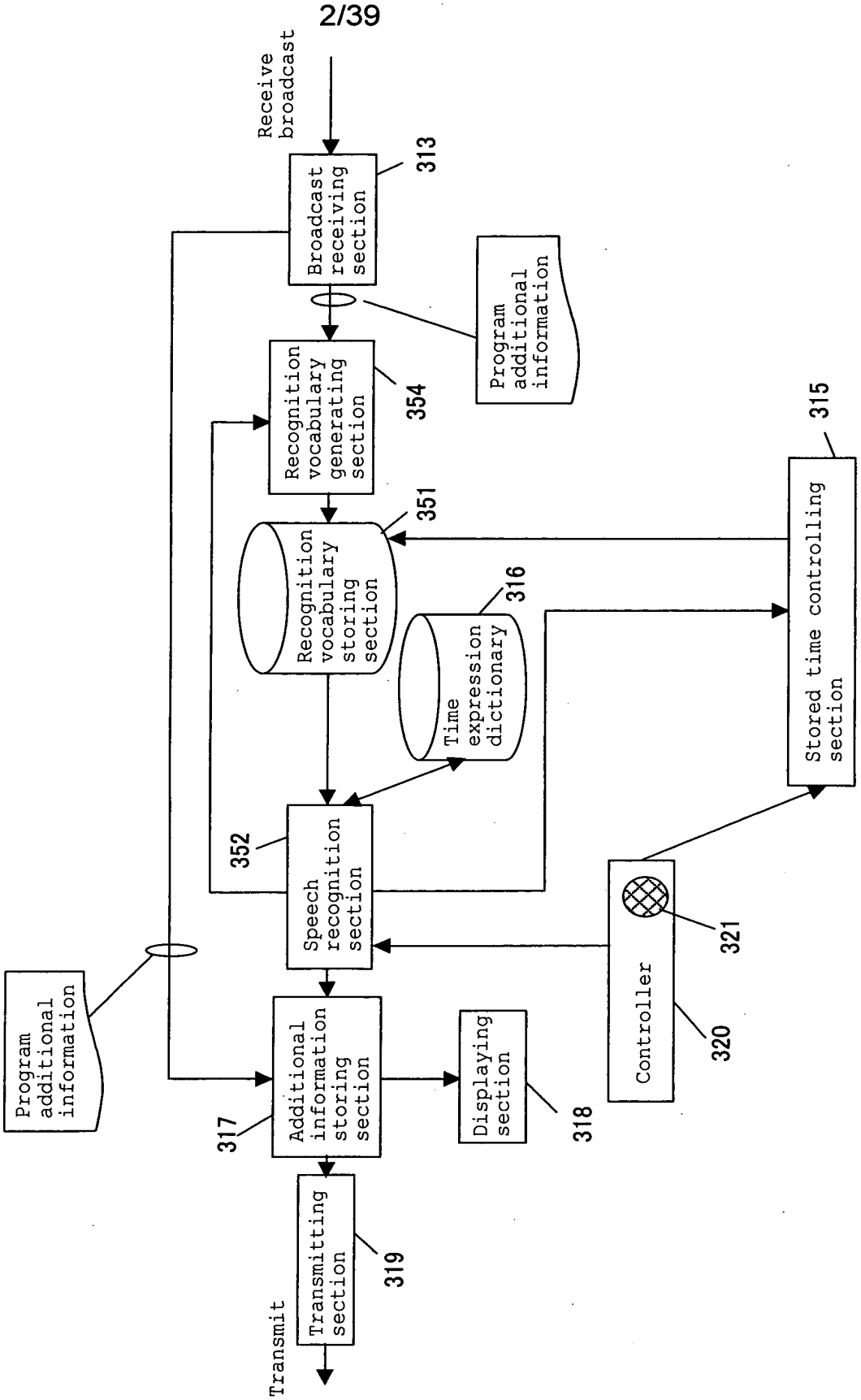

Receive broadcast
Broadcast receiving section
313
Program additional information
Recognition vocabulary generating section
354
Recognition vocabulary storing section
351
Time expression dictionary
316
Stored time controlling section
315
Speech recognition section
352
Program additional information
Additional information storing section
317
Displaying section
318
Controller
320
321
Transmitting section
319
Transmit

Fig.2

10/542409

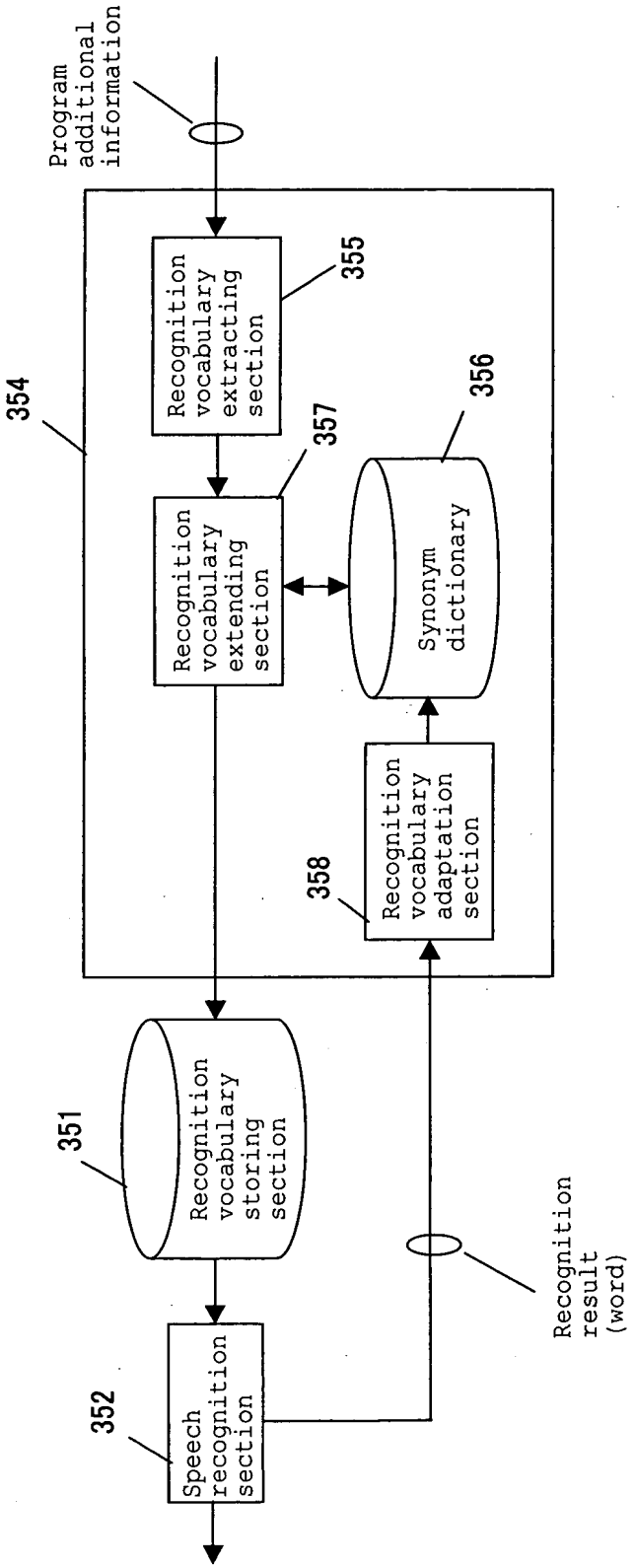


Fig.3

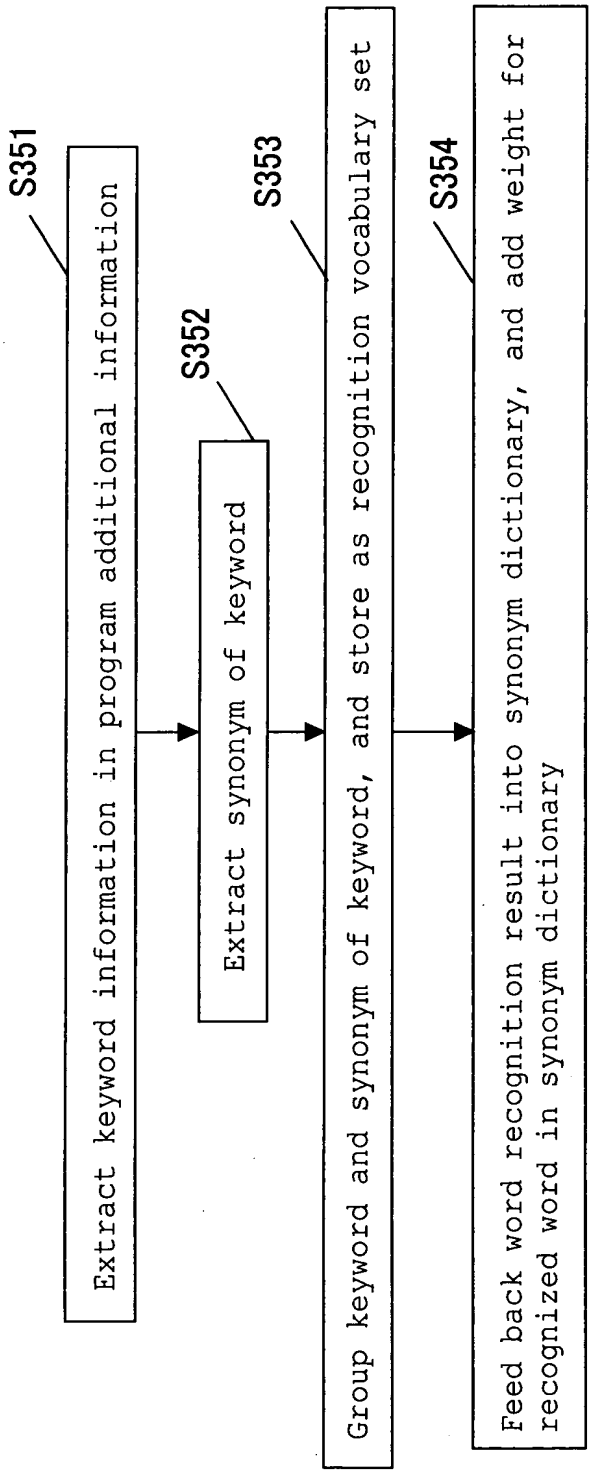

S351
Extract keyword information in program additional information
S352
Extract synonym of keyword
S353
Group keyword and synonym of keyword, and store as recognition vocabulary set
S354
Feed back word recognition result into synonym dictionary, and add weight for recognized word in synonym dictionary

Fig.4

10/542409

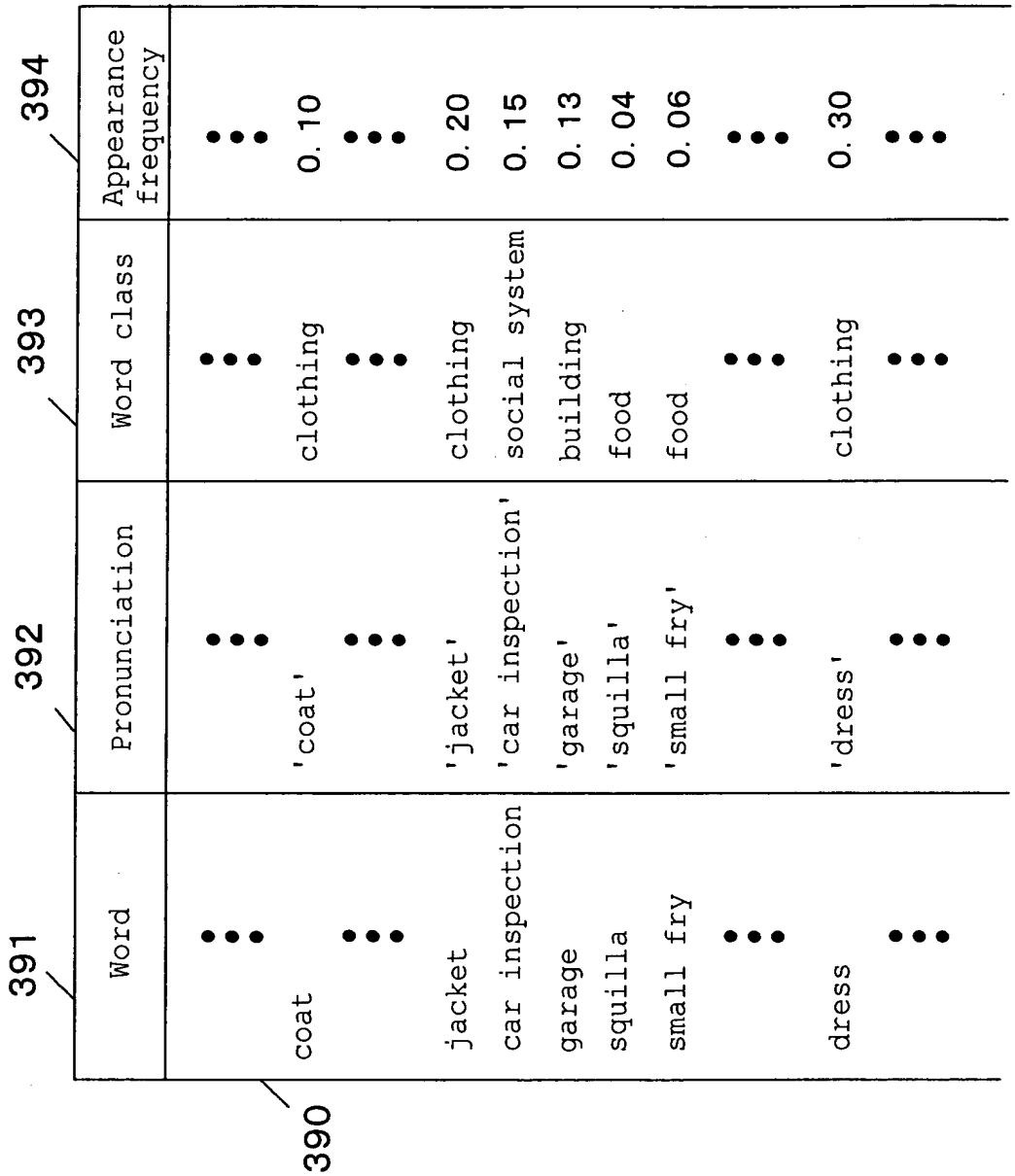


Fig.5

10/542409

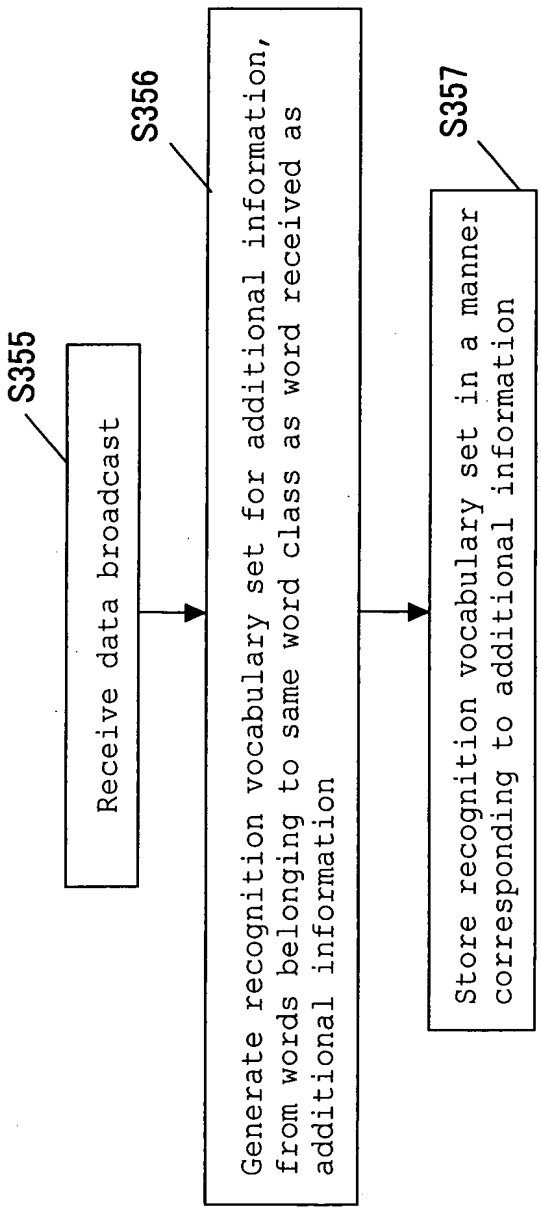


Fig.6

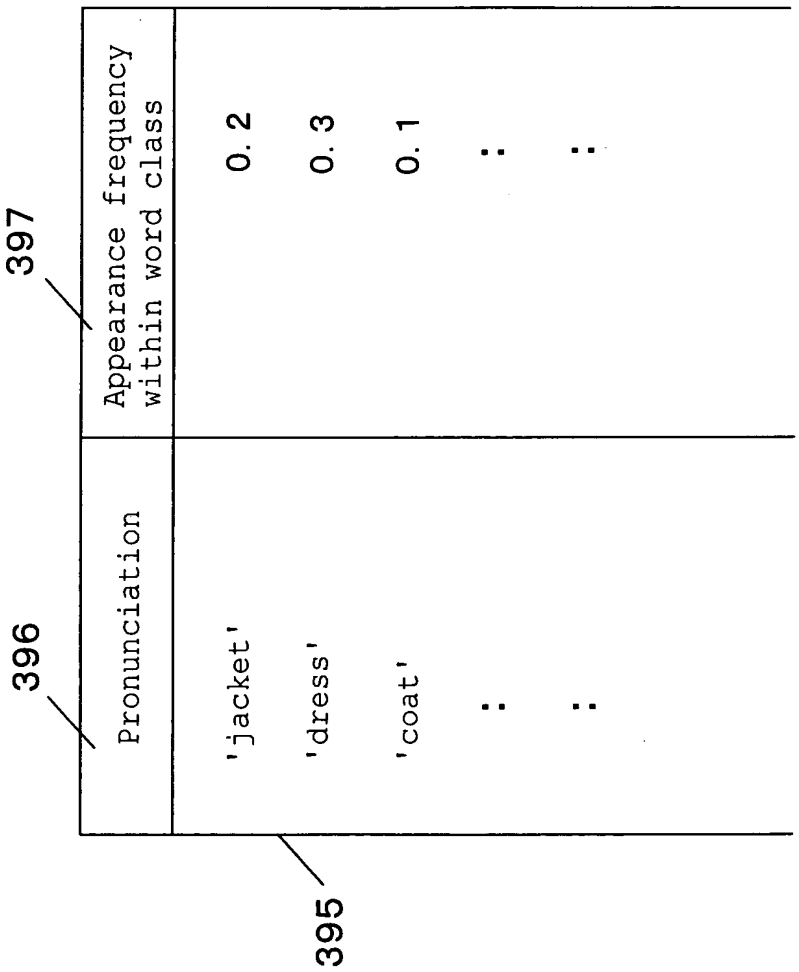

397
396
395
Pronunciation
Appearance frequency within word class
'jacket'
'dress'
'coat'
0.2
0.3
0.1

Fig.7

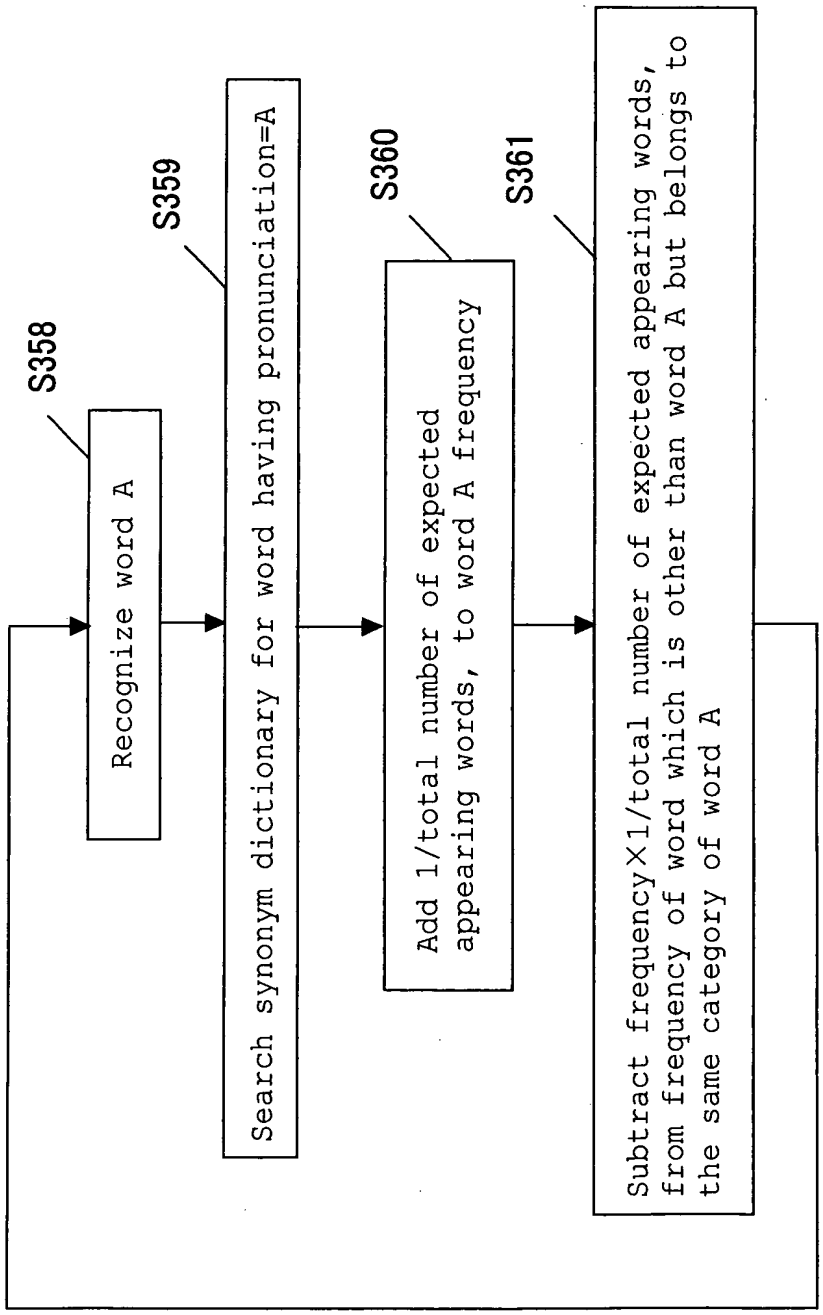

S358
Recognize word A
S359
Search synonym dictionary for word having pronunciation=A
S360
Add 1/total number of expected appearing words, to word A frequency
S361
Subtract frequency×1/total number of expected appearing words, from frequency of word which is other than word A but belongs to the same category of word A

Fig.8

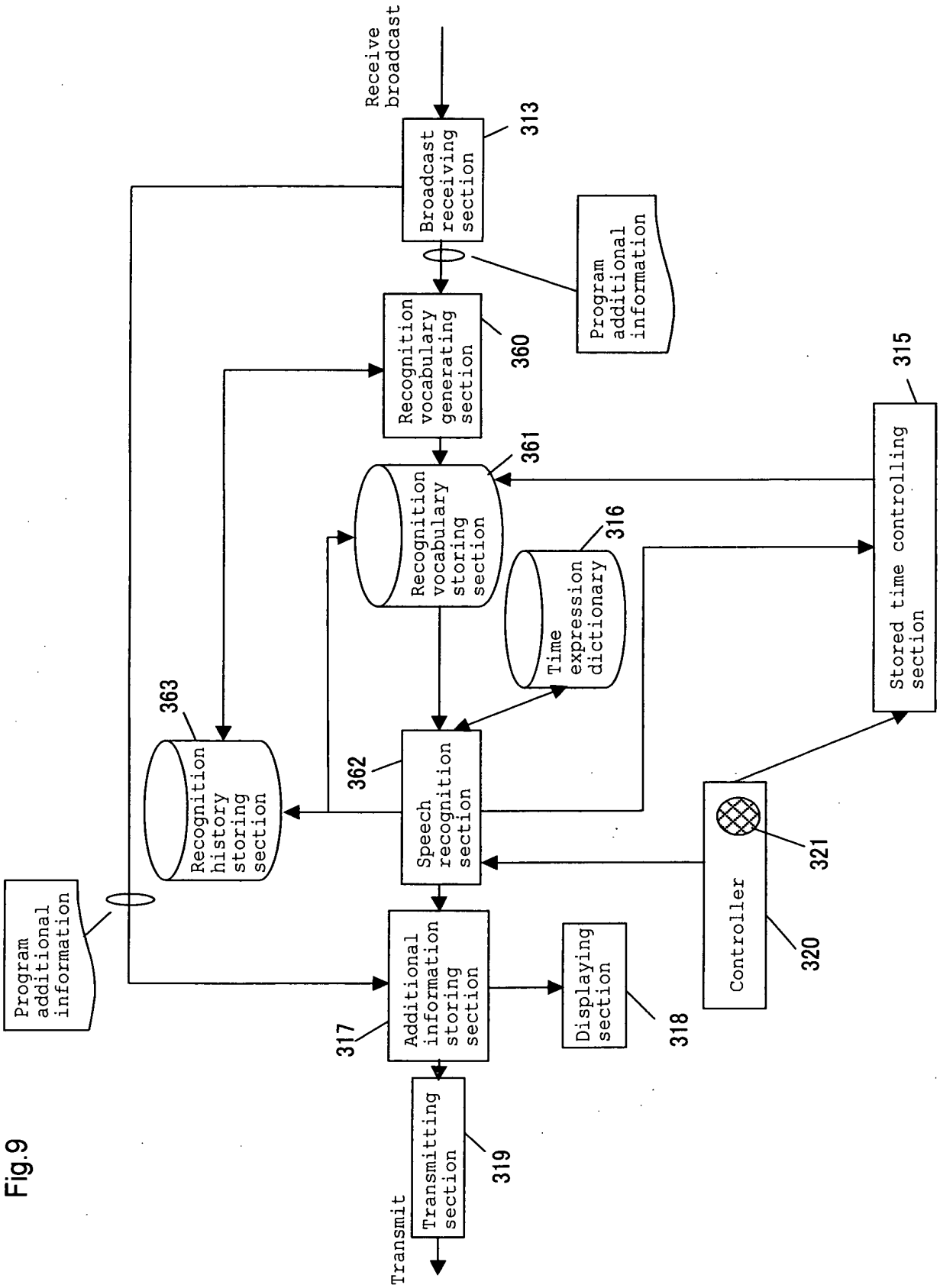

Receive broadcast
Broadcast receiving section
313
Program additional information
Recognition vocabulary generating section
360
315
Recognition vocabulary storing section
361
316
Time expression dictionary
Stored time controlling section
363
Recognition history storing section
362
Speech recognition section
Program additional information
Controller
321
320
Additional information storing section
317
Displaying section
318
Transmitting section
319
Transmit

Fig.9

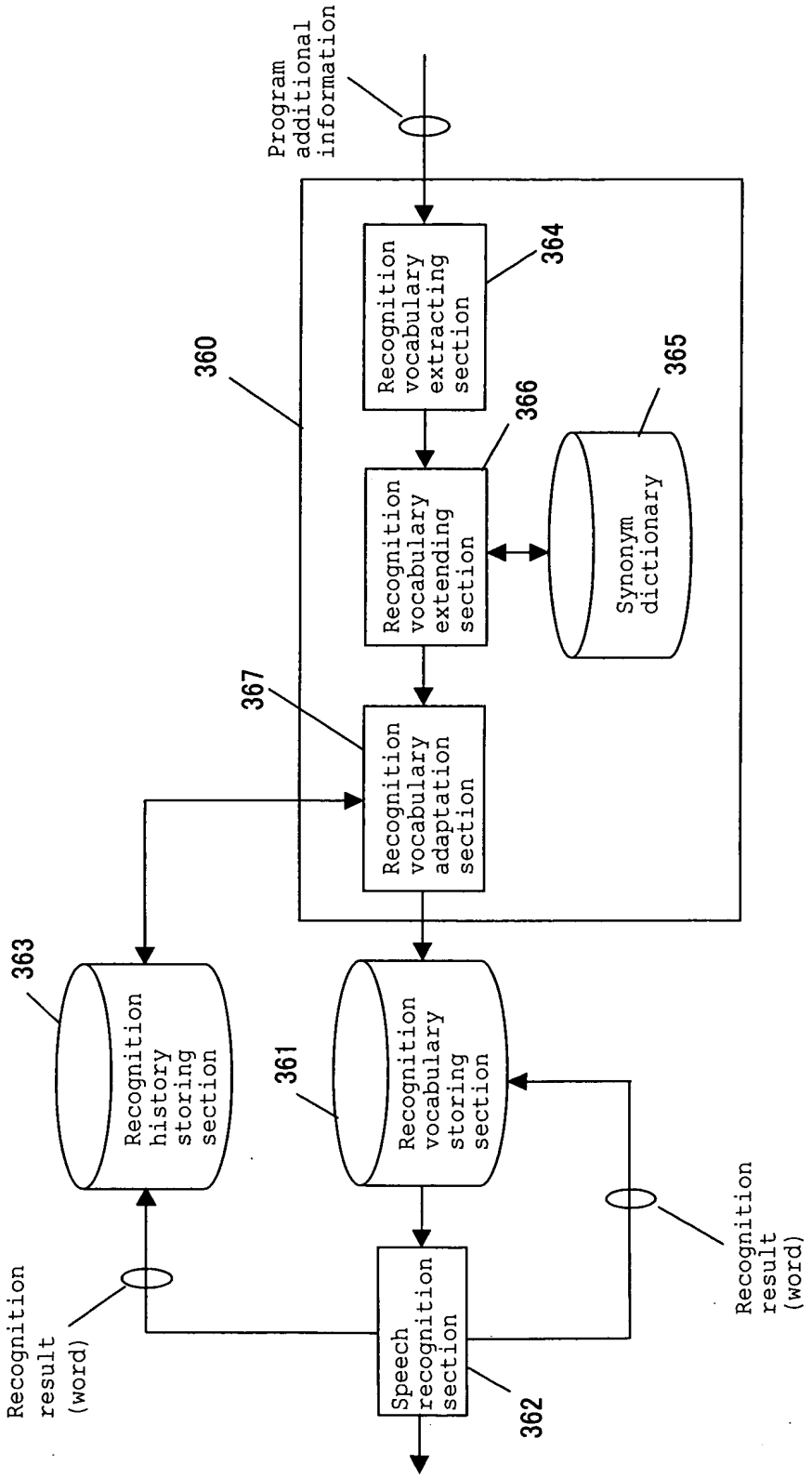

Program additional information
360
Recognition vocabulary extracting section
364
Recognition vocabulary extending section
366
Synonym dictionary
365
Recognition vocabulary adaptation section
367
Recognition history storing section
363
Recognition vocabulary storing section
361
Speech recognition section
362
Recognition result (word)
Recognition result (word)

Fig.10

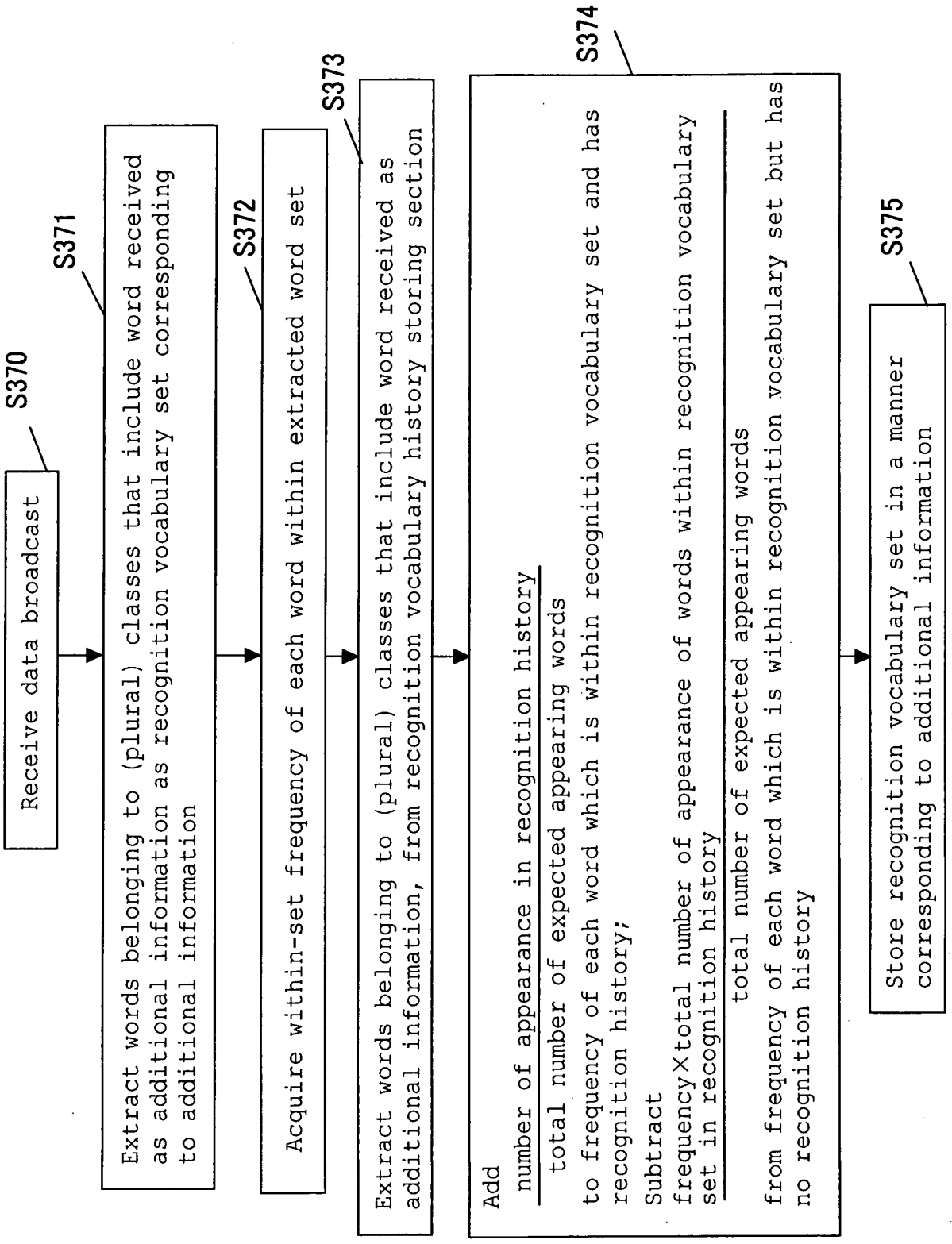
10/542409

Fig.11

400

401 Word	461 Pronunciation	402 Word class 1	403 Word class 2	404 ... Word class n	405 Appearance frequency
jacket	'jacket'	clothing	fashion		0.020
car inspection	'car inspection'	social system	transportation		0.015
garage	'garage'	building	transportation		0.013
squilla	'squilla'	food	shellfish		0.004
small fry	'small fry'	food	fish		0.006

Fig.12


S370
Receive data broadcast
S371
Extract words belonging to (plural) classes that include word received as additional information as recognition vocabulary set corresponding to additional information
S372
Acquire within-set frequency of each word within extracted word set
S373
Extract words belonging to (plural) classes that include word received as additional information, from recognition vocabulary history storing section
S374
Add
number of appearance in recognition history
total number of expected appearing words
to frequency of each word which is within recognition vocabulary set and has recognition history;
Subtract
frequency×total number of appearance of words within recognition vocabulary set in recognition history
total number of expected appearing words
from frequency of each word which is within recognition vocabulary set but has no recognition history
S375
Store recognition vocabulary set in a manner corresponding to additional information

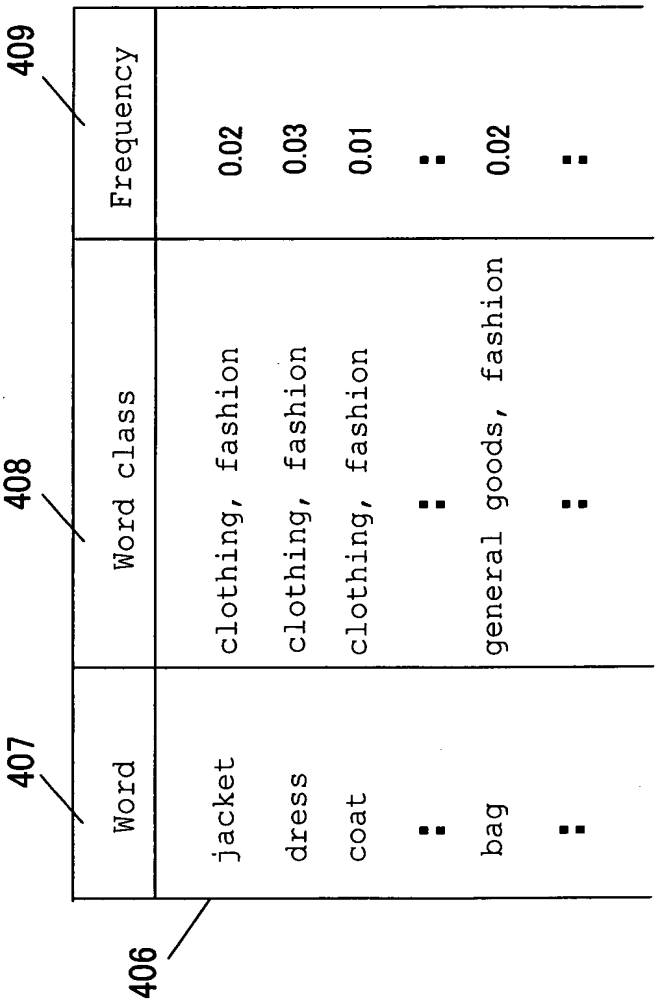

406
407
Word
408
Word class
409
Frequency
jacket
clothing, fashion
0.02
dress
clothing, fashion
0.03
coat
clothing, fashion
0.01
bag
general goods, fashion
0.02

Fig.13

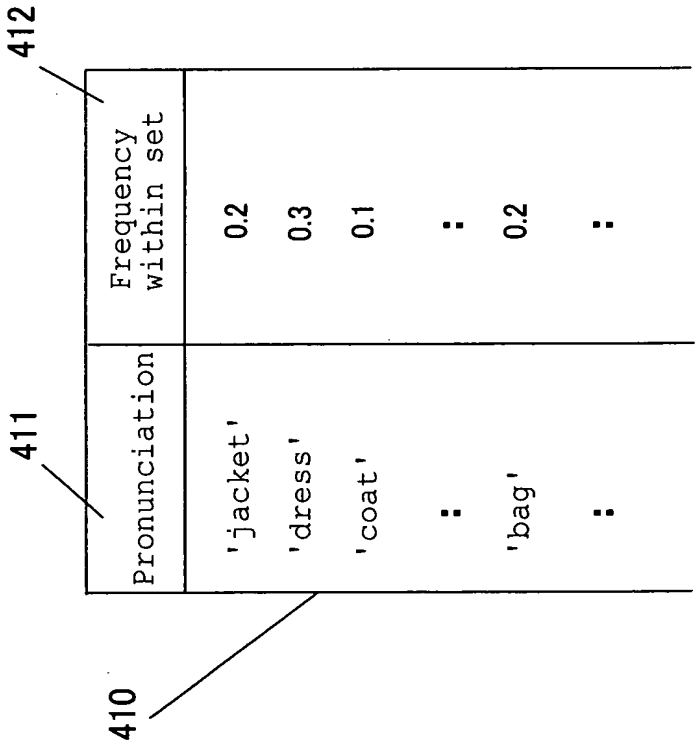

412
411
410
Pronunciation
Frequency within set
'jacket'
'dress'
'coat'
'bag'
0.2
0.3
0.1
0.2

Fig.14

10/542409

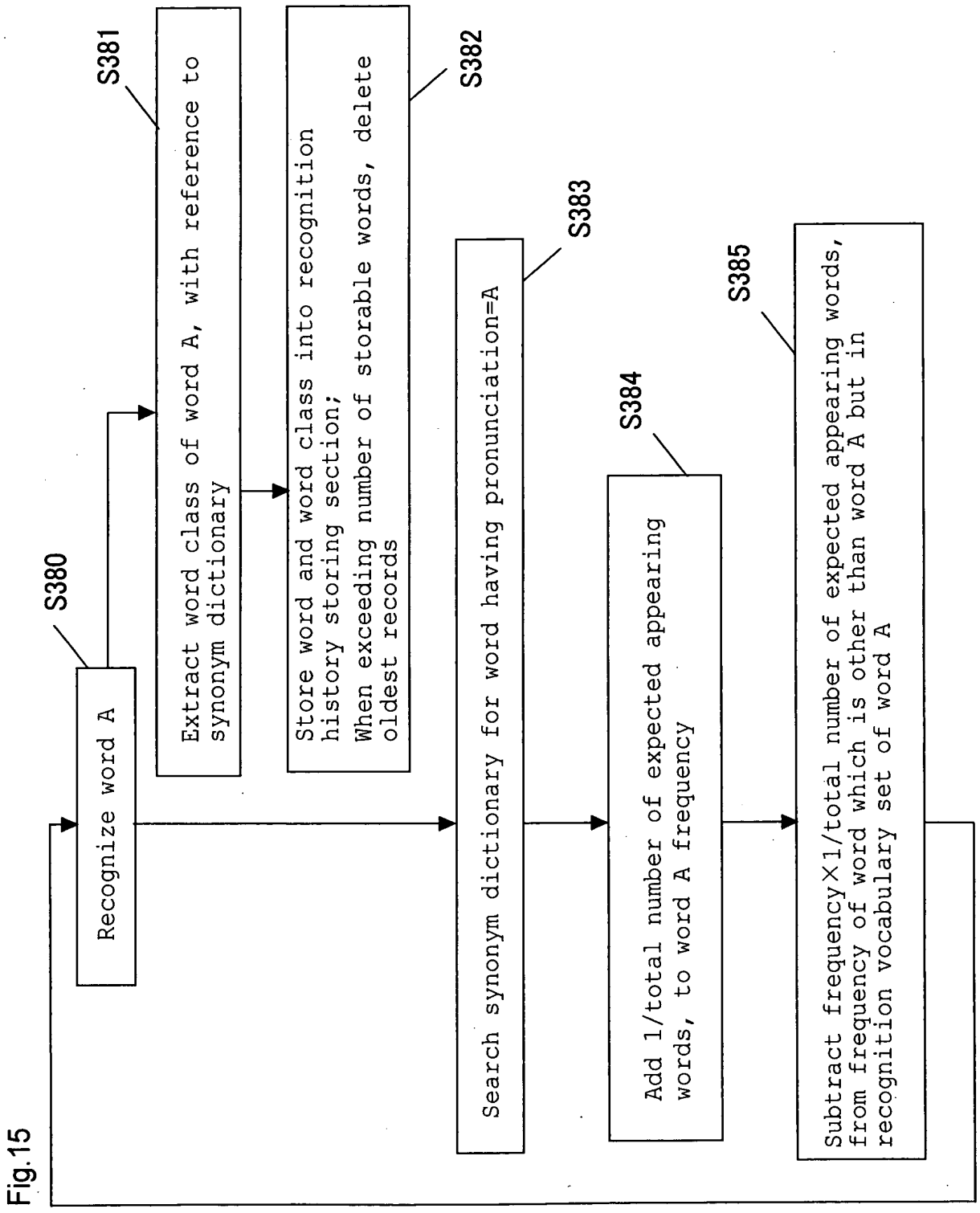


Fig.15

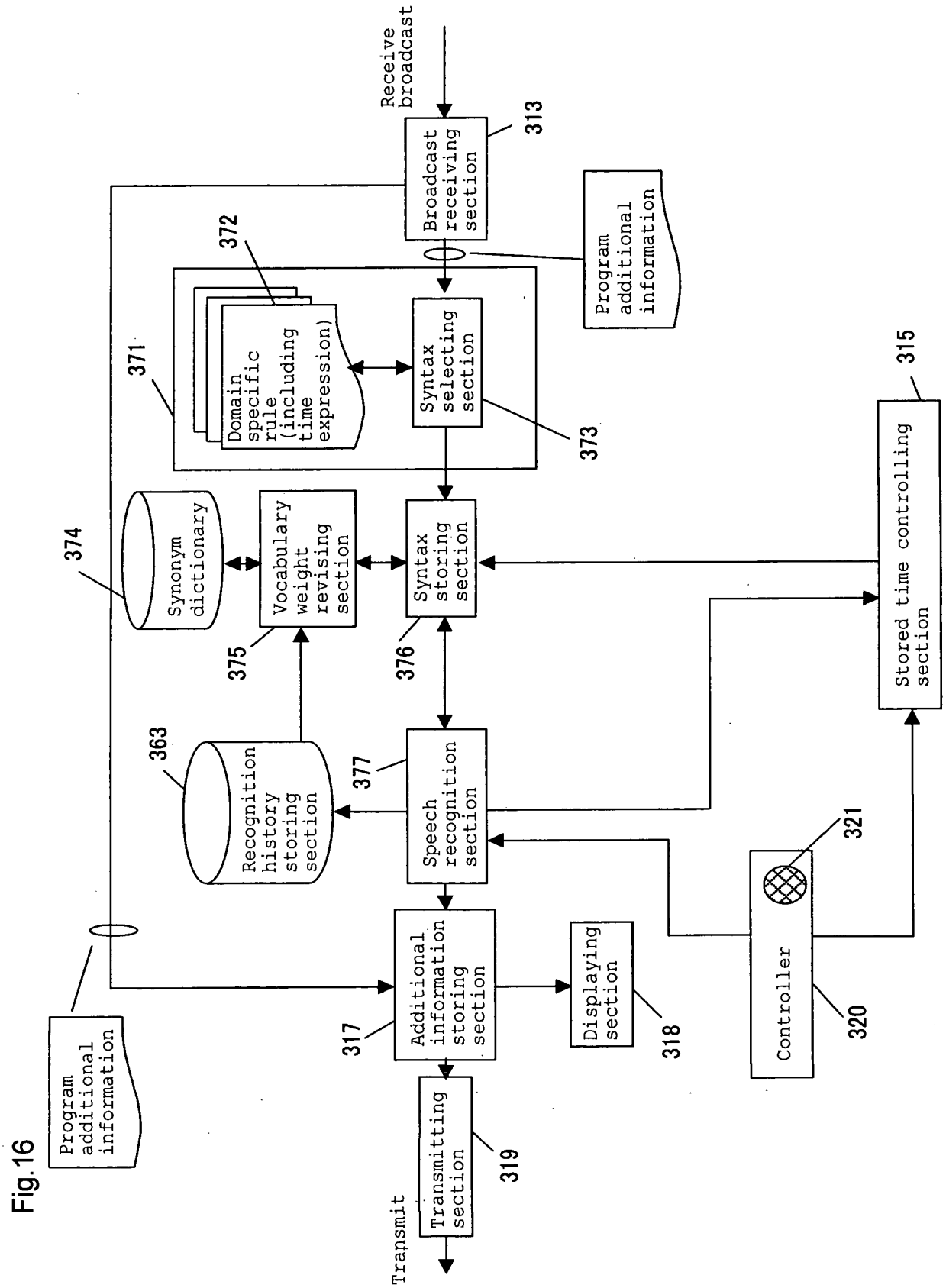

Receive broadcast
Broadcast receiving section
313
Program additional information
372
371
Domain specific rule (including time expression)
Syntax selecting section
373
315
Stored time controlling section
374
Synonym dictionary
Vocabulary weight revising section
Syntax storing section
375
376
363
Recognition history storing section
377
Speech recognition section
321
Additional information storing section
317
Displaying section
318
Controller
320
Program additional information
Transmitting section
319
Transmit

Fig.16

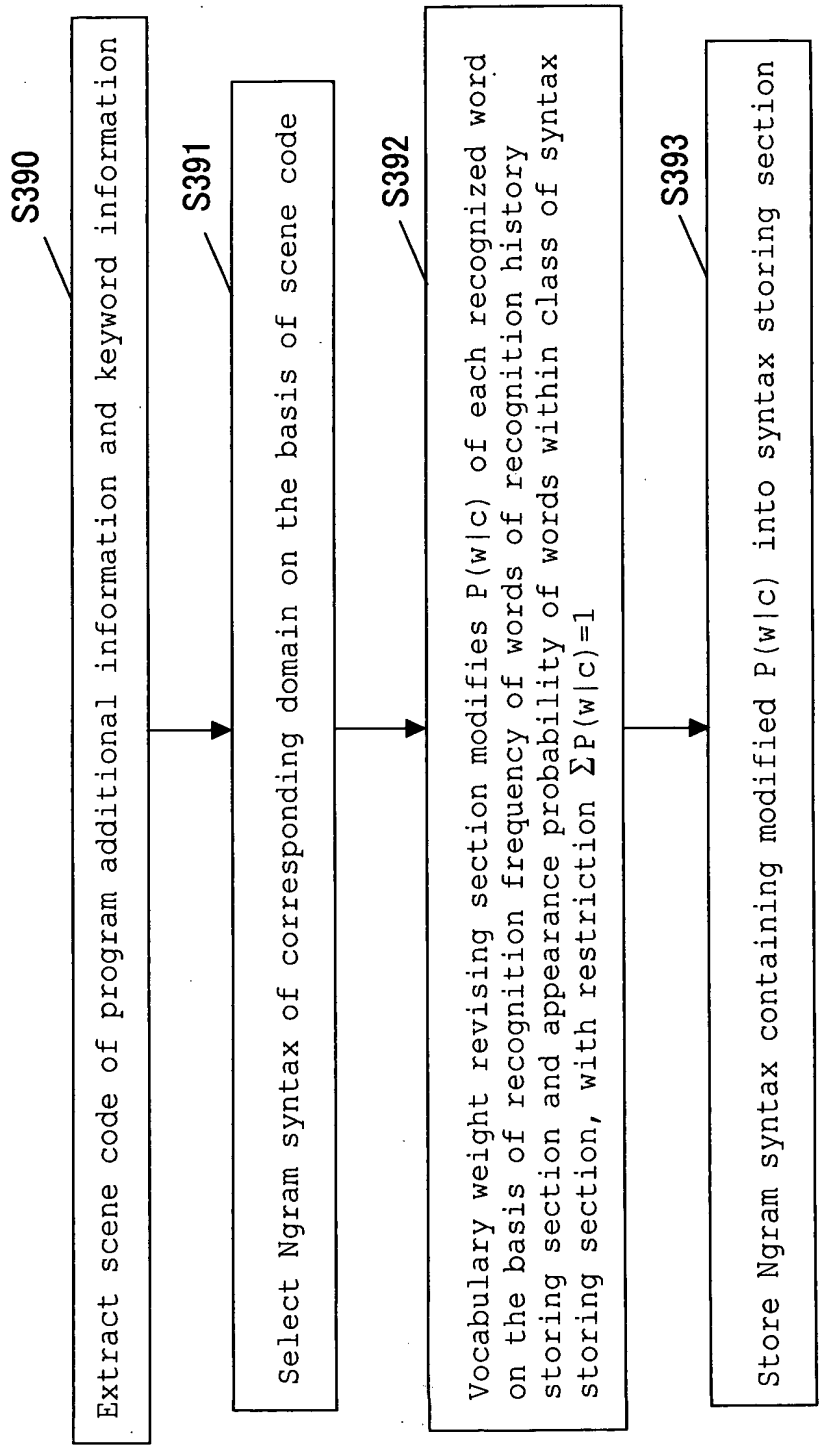

S390
Extract scene code of program additional information and keyword information
S391
Select Ngram syntax of corresponding domain on the basis of scene code
S392
Vocabulary weight revising section modifies P(w|c) of each recognized word on the basis of recognition frequency of words of recognition history storing section and appearance probability of words within class of syntax storing section, with restriction ΣP(w|c)=1
S393
Store Ngram syntax containing modified P(w|c) into syntax storing section

Fig.17

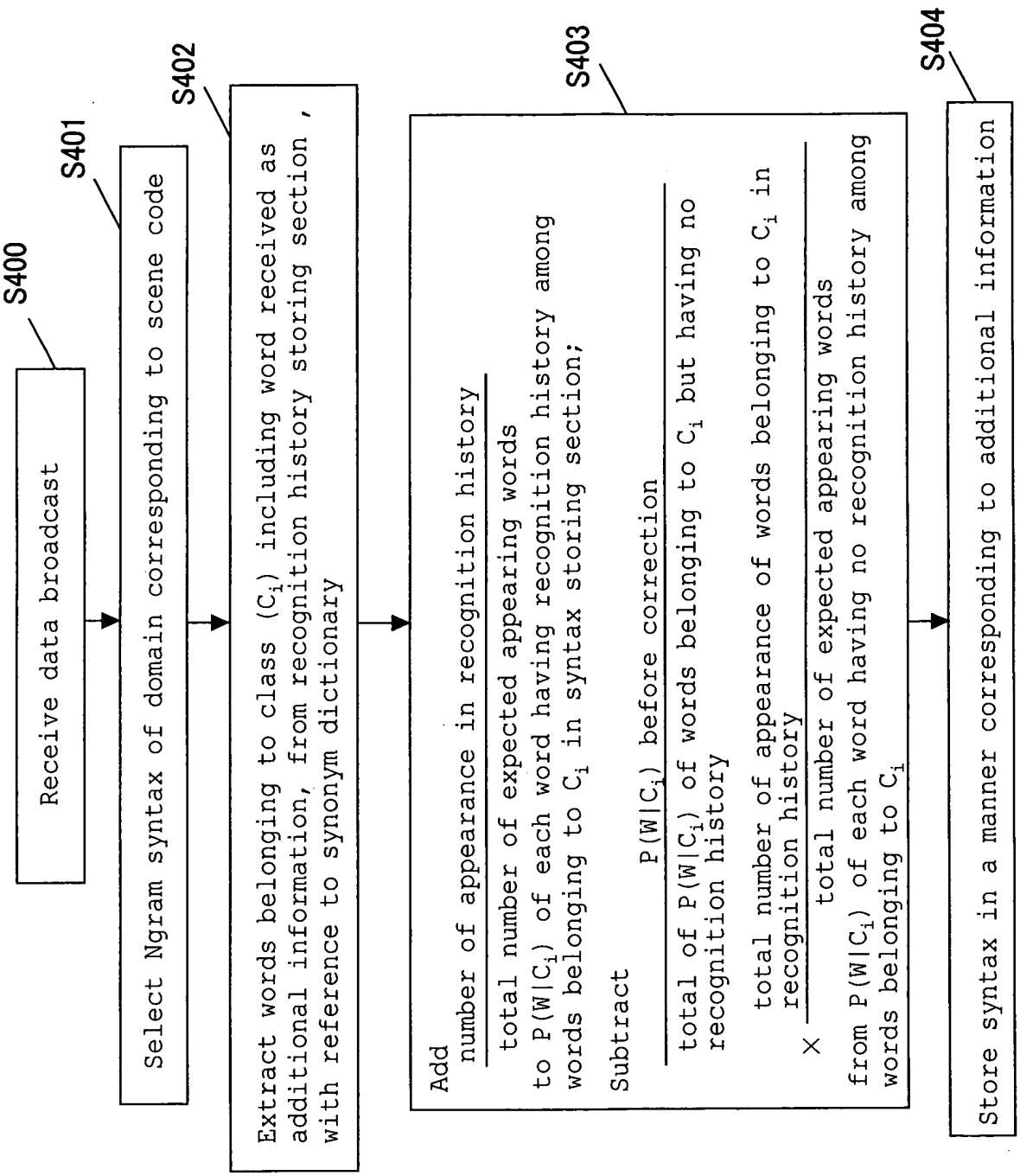

S400
Receive data broadcast
S401
Select Ngram syntax of domain corresponding to scene code
S402
Extract words belonging to class (C_i) including word received as additional information, from recognition history storing section , with reference to synonym dictionary
S403
Add
number of appearance in recognition history / total number of expected appearing words
to $P(W|C_i)$ of each word having recognition history among words belonging to C_i in syntax storing section;
Subtract
$P(W|C_i)$ before correction / total of $P(W|C_i)$ of words belonging to C_i but having no recognition history
× total number of appearance of words belonging to C_i in recognition history / total number of expected appearing words
from $P(W|C_i)$ of each word having no recognition history among words belonging to C_i
S404
Store syntax in a manner corresponding to additional information

Fig.18

Fig.19

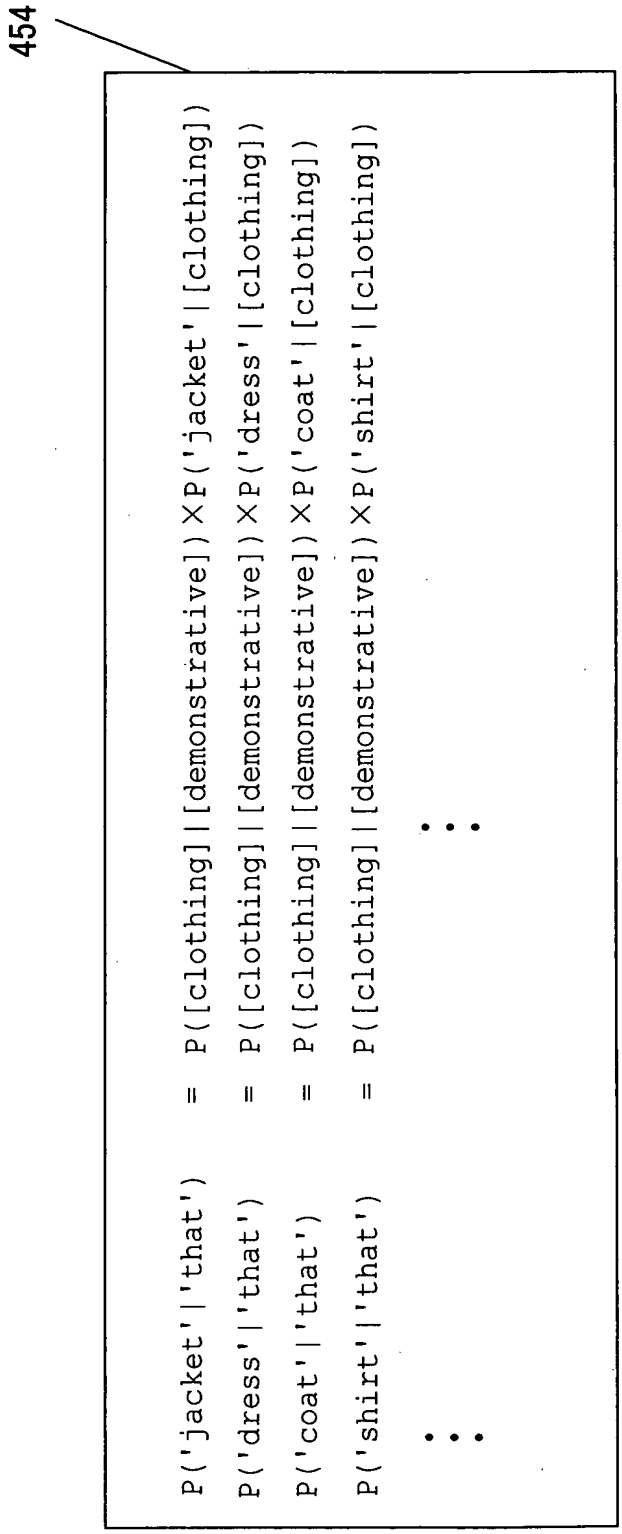
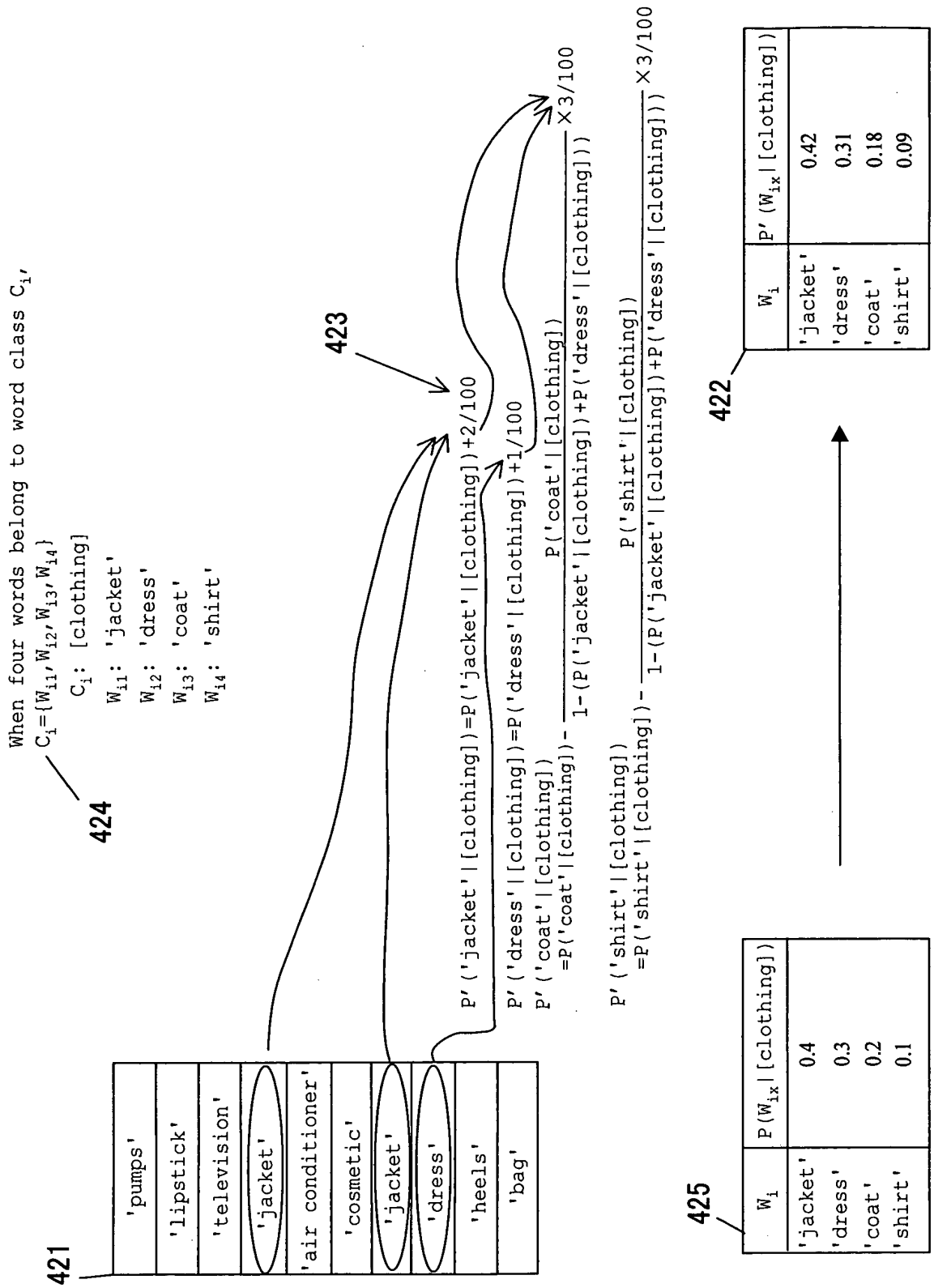

454
P('jacket'|'that') = P([clothing]|[demonstrative])×P('jacket'|[clothing])
P('dress'|'that') = P([clothing]|[demonstrative])×P('dress'|[clothing])
P('coat'|'that') = P([clothing]|[demonstrative])×P('coat'|[clothing])
P('shirt'|'that') = P([clothing]|[demonstrative])×P('shirt'|[clothing])
• • •
• • •

Fig.20


When four words belong to word class C_i,
$C_i=\{W_{i1},W_{i2},W_{i3},W_{i4}\}$
C_i: [clothing]
W_{i1}: 'jacket'
W_{i2}: 'dress'
W_{i3}: 'coat'
W_{i4}: 'shirt'
424
421
'pumps'
'lipstick'
'television'
'jacket'
'air conditioner'
'cosmetic'
'jacket'
'dress'
'heels'
'bag'
423
P'('jacket'|[clothing])=P('jacket'|[clothing])+2/100
P'('dress'|[clothing])=P('dress'|[clothing])+1/100
P'('coat'|[clothing])=P('coat'|[clothing])-P('coat'|[clothing])/(1-(P('jacket'|[clothing])+P('dress'|[clothing])))×3/100
P'('shirt'|[clothing])=P('shirt'|[clothing])-P('shirt'|[clothing])/(1-(P('jacket'|[clothing])+P('dress'|[clothing])))×3/100
422
W_i | $P'(W_{ix}|[clothing])$
'jacket' | 0.42
'dress' | 0.31
'coat' | 0.18
'shirt' | 0.09
425
W_i | $P(W_{ix}|[clothing])$
'jacket' | 0.4
'dress' | 0.3
'coat' | 0.2
'shirt' | 0.1

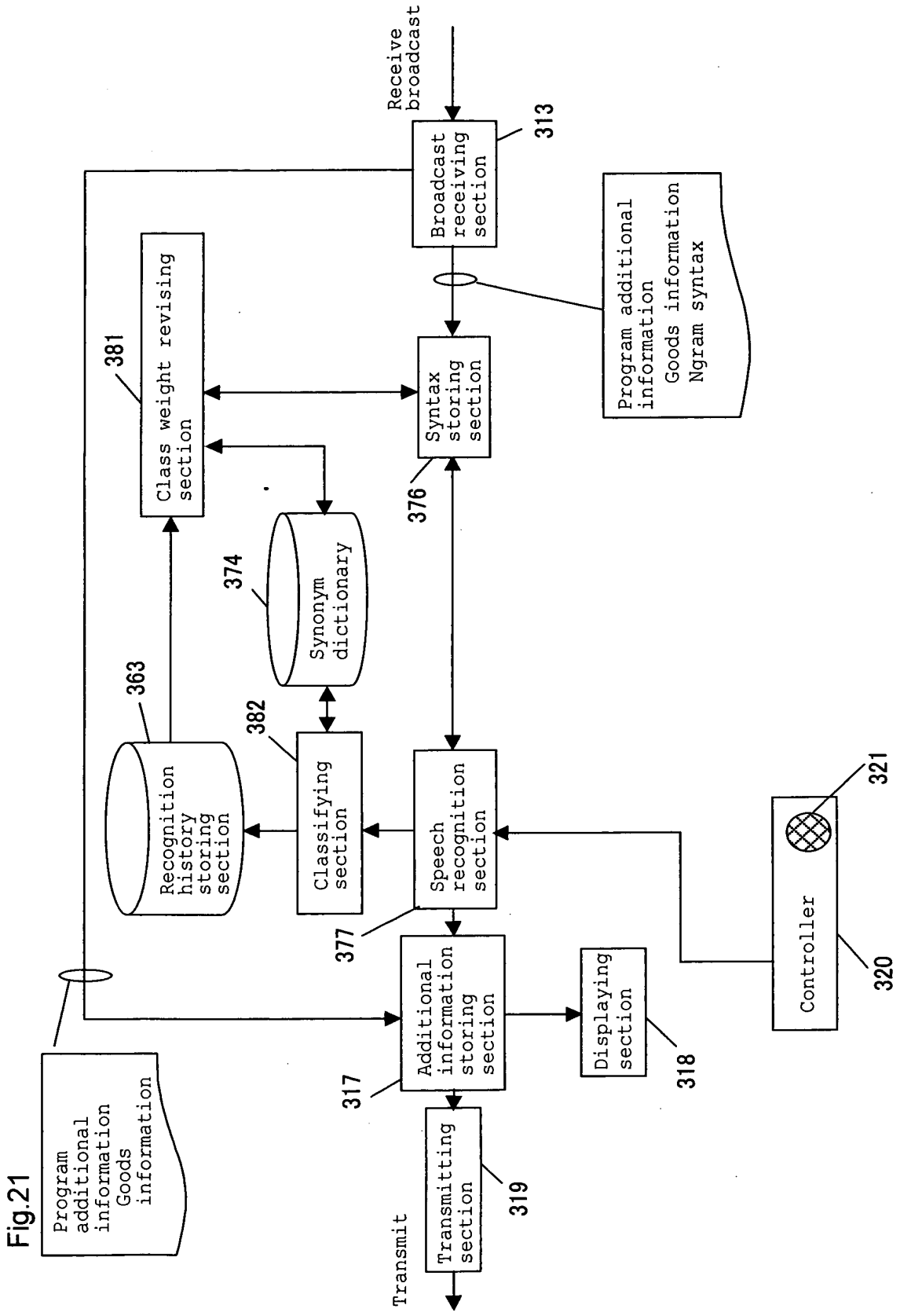

Receive broadcast
313
Broadcast receiving section
Program additional information
Goods information
Ngram syntax
381
Class weight revising section
Syntax storing section
376
374
Synonym dictionary
363
382
Recognition history storing section
Classifying section
Speech recognition section
321
Controller
320
377
Additional information storing section
Displaying section
318
317
Transmitting section
319
Transmit
Program additional information
Goods information

Fig.21

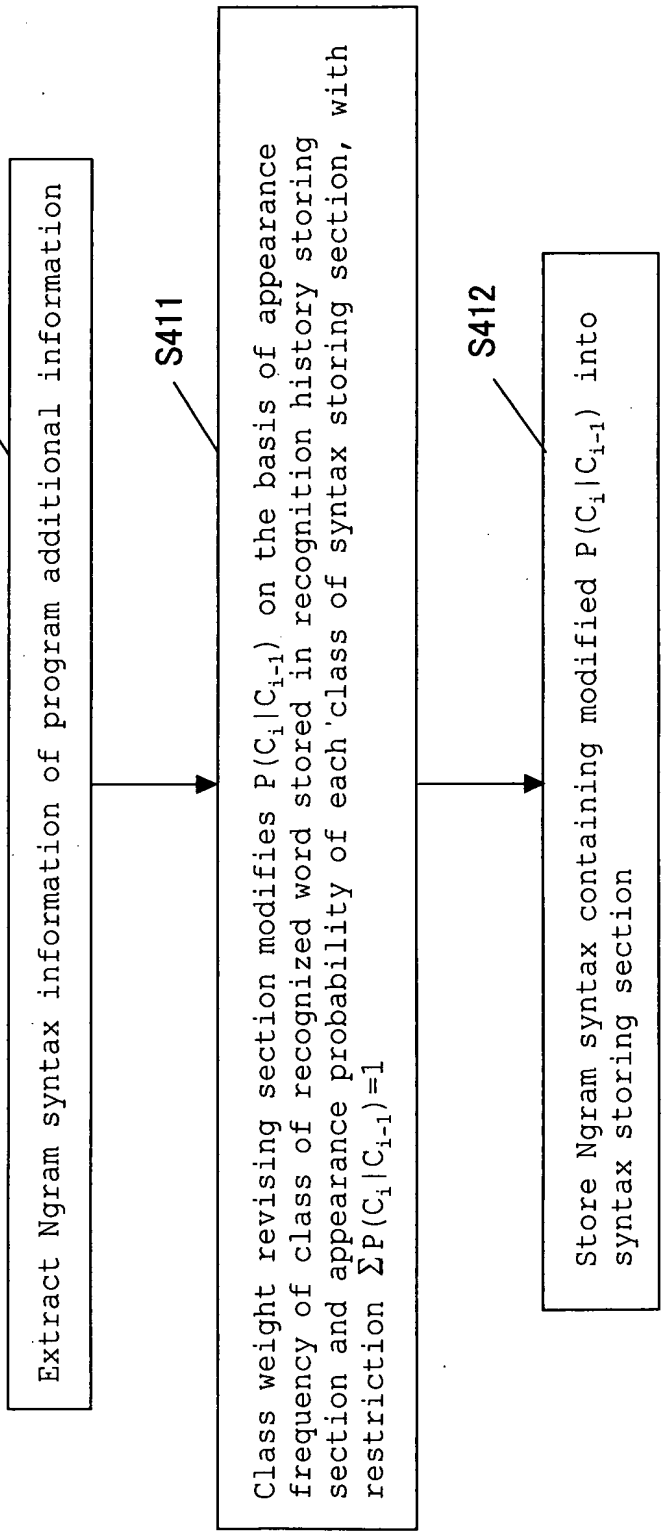

S410
Extract Ngram syntax information of program additional information
S411
Class weight revising section modifies $P(C_i|C_{i-1})$ on the basis of appearance frequency of class of recognized word stored in recognition history storing section and appearance probability of each 'class of syntax storing section, with restriction $\sum P(C_i|C_{i-1})=1$
S412
Store Ngram syntax containing modified $P(C_i|C_{i-1})$ into syntax storing section

Fig.22

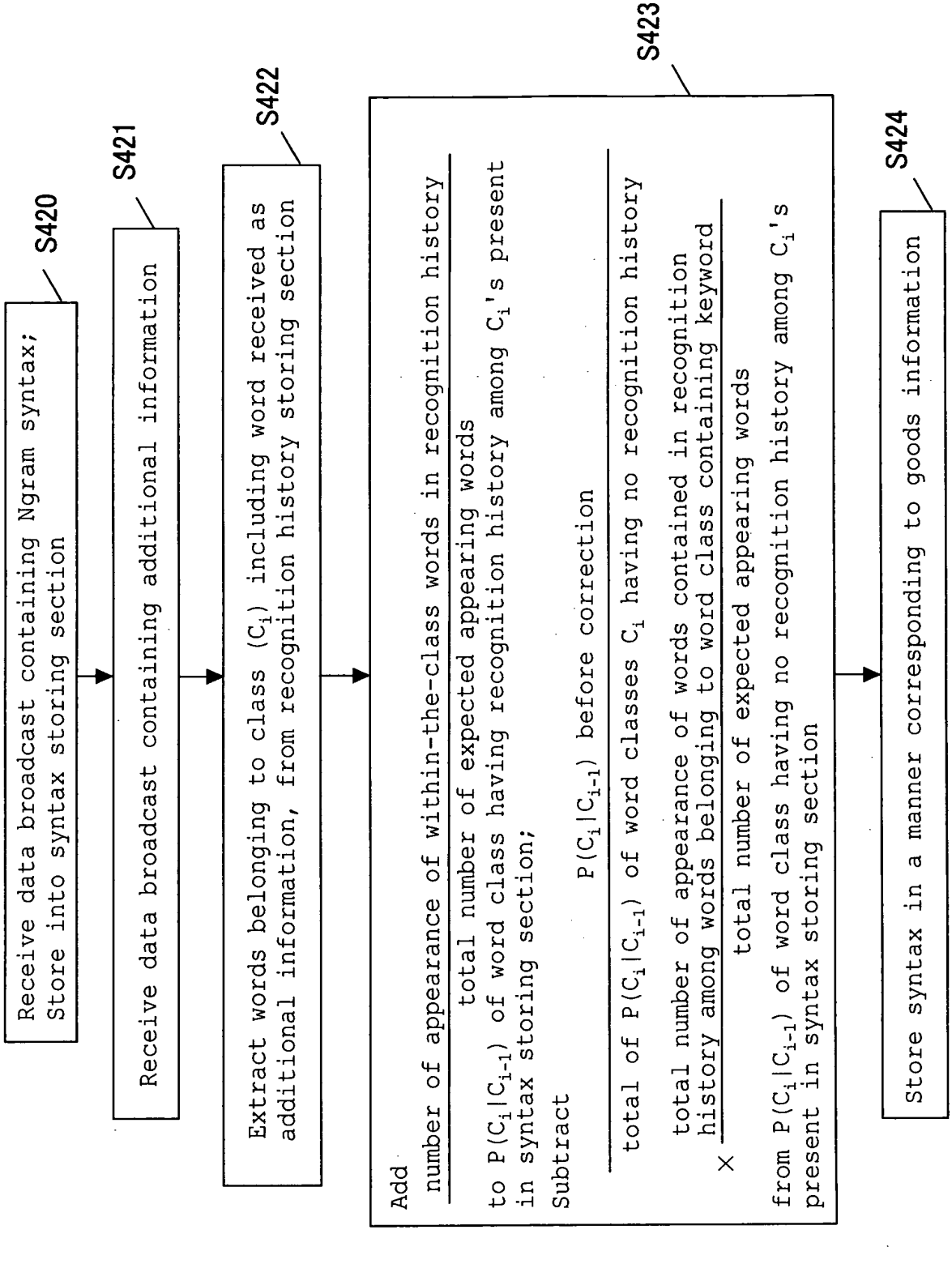

S420
Receive data broadcast containing Ngram syntax; Store into syntax storing section
S421
Receive data broadcast containing additional information
S422
Extract words belonging to class (C_i) including word received as additional information, from recognition history storing section
S423
Add
number of appearance of within-the-class words in recognition history
total number of expected appearing words
to $P(C_i|C_{i-1})$ of word class having recognition history among C_i's present in syntax storing section;
Subtract
$P(C_i|C_{i-1})$ before correction
total of $P(C_i|C_{i-1})$ of word classes C_i having no recognition history
total number of appearance of words contained in recognition history among words belonging to word class containing keyword
×
total number of expected appearing words
from $P(C_i|C_{i-1})$ of word class having no recognition history among C_i's present in syntax storing section
S424
Store syntax in a manner corresponding to goods information

Fig.23

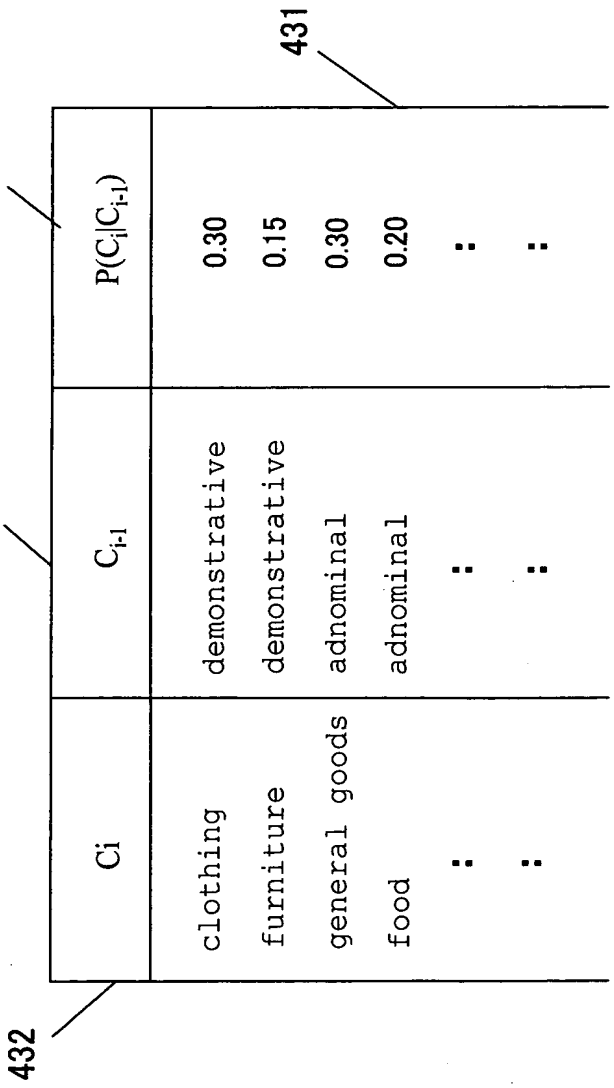
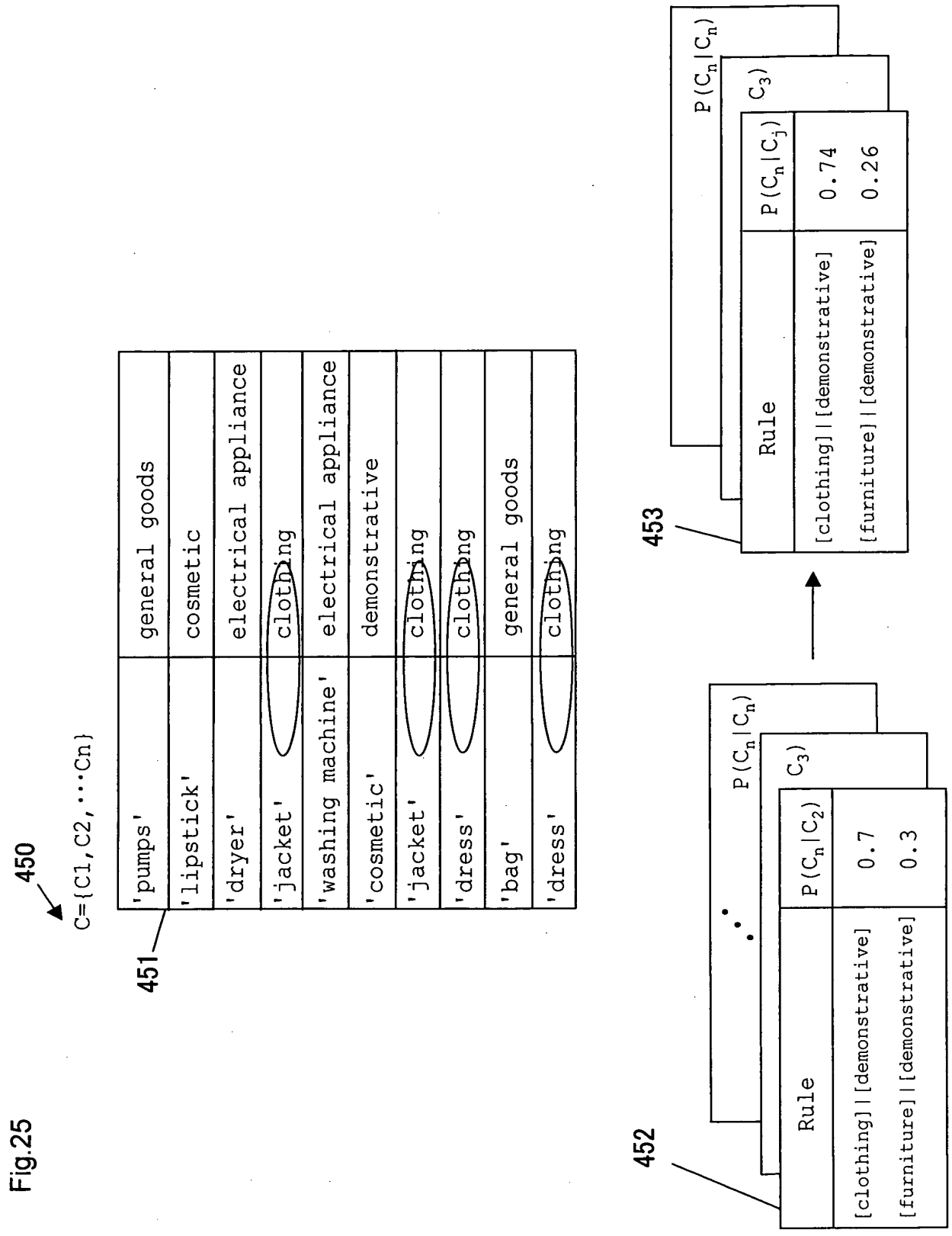

431
432
433
434
Ci
C_{i-1}
$P(C_i|C_{i-1})$
clothing
furniture
general goods
food
demonstrative
demonstrative
adnominal
adnominal
0.30
0.15
0.30
0.20

Fig.24

Fig.25


450
C={C1,C2,···Cn}
451
'pumps' general goods
'lipstick' cosmetic
'dryer' electrical appliance
'jacket' clothing
'washing machine' electrical appliance
'cosmetic' demonstrative
'jacket' clothing
'dress' clothing
'bag' general goods
'dress' clothing
452
Rule
P(Cn|C2)
C3)
P(Cn|Cn)
[clothing]|[demonstrative] 0.7
[furniture]|[demonstrative] 0.3
453
Rule
P(Cn|Cj)
C3)
P(Cn|Cn)
[clothing]|[demonstrative] 0.74
[furniture]|[demonstrative] 0.26

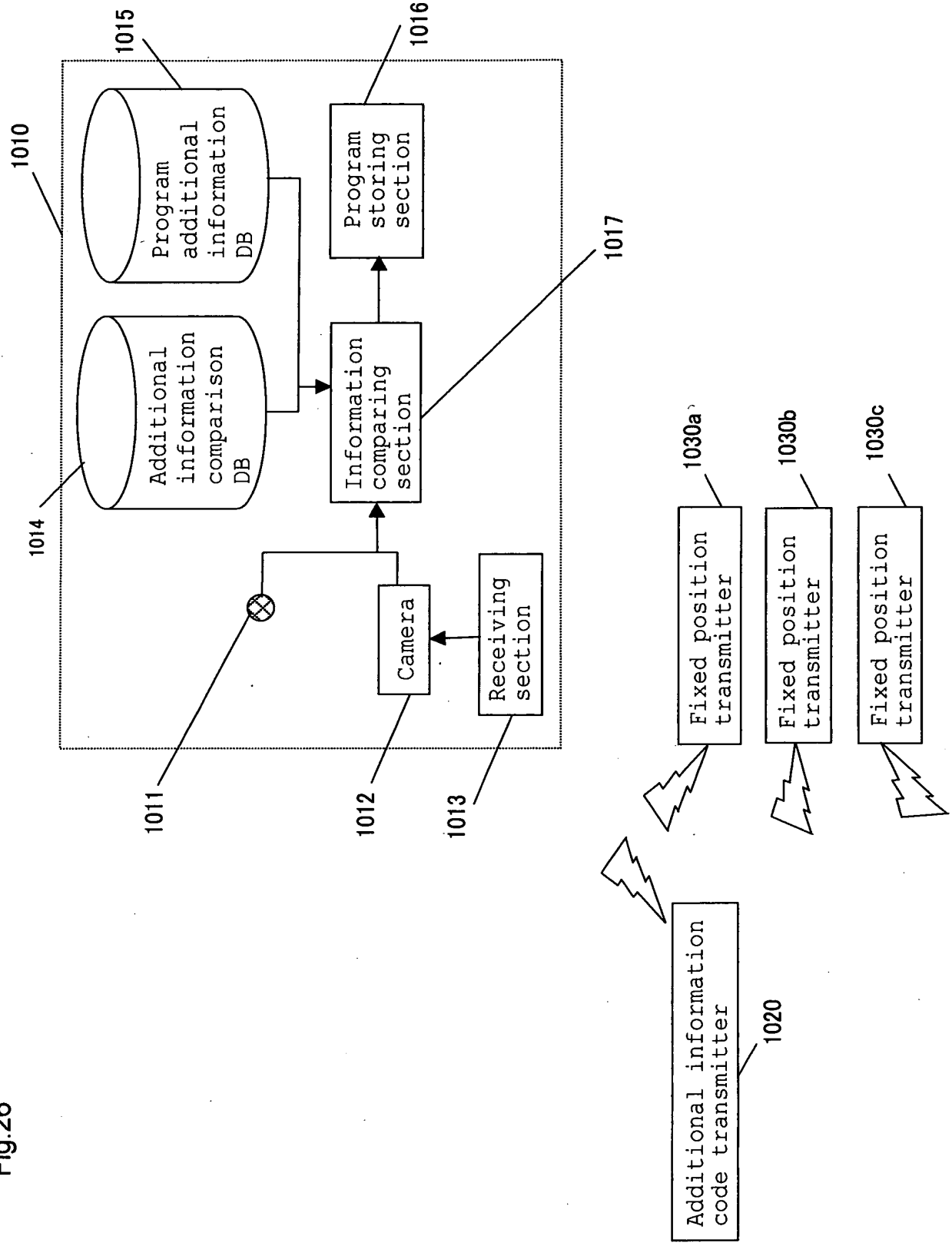

1010
1011
1012
1013
1014
1015
1016
1017
1020
1030a
1030b
1030c
Program additional information DB
Additional information comparison DB
Program storing section
Information comparing section
Camera
Receiving section
Fixed position transmitter
Fixed position transmitter
Fixed position transmitter
Additional information code transmitter

Fig.26

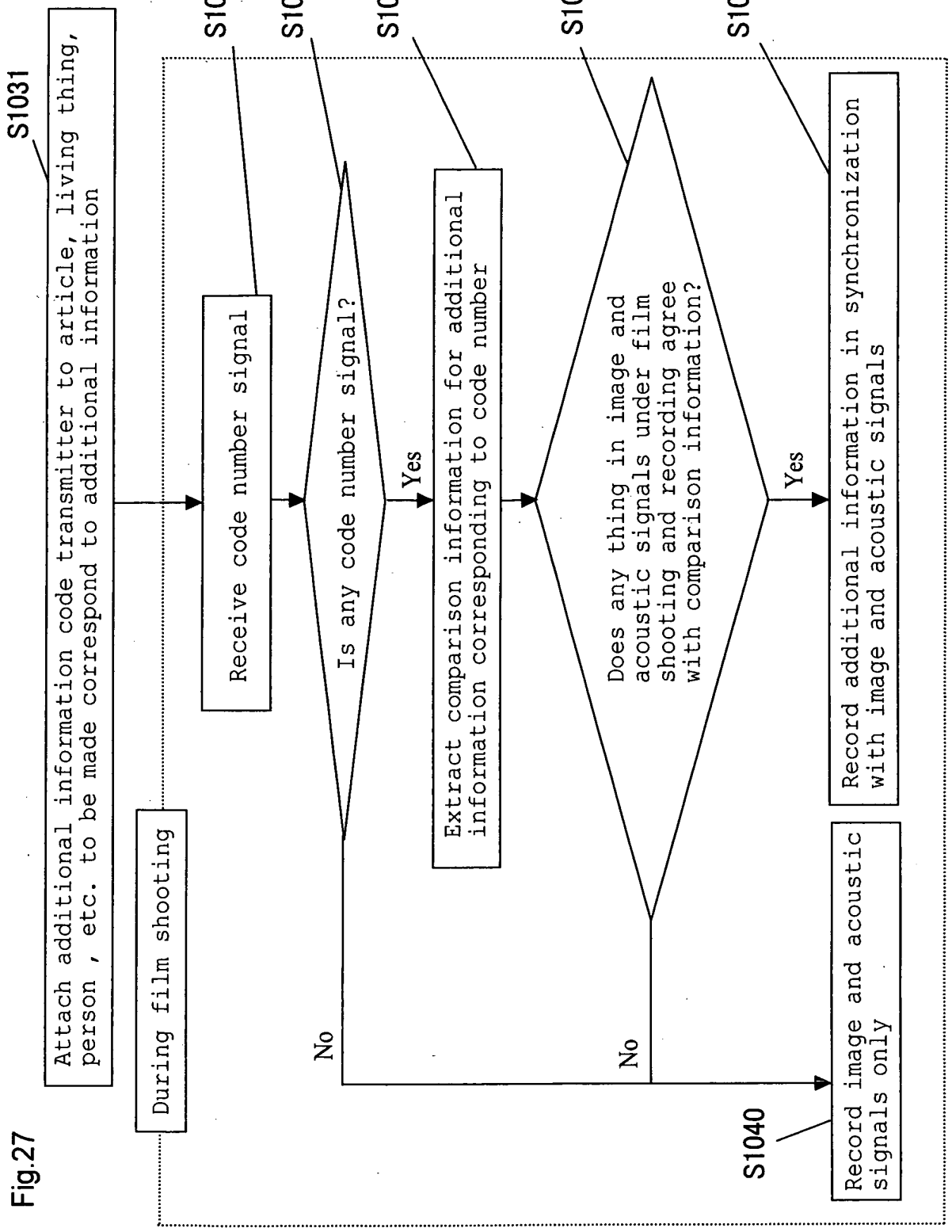

S1031
Attach additional information code transmitter to article, living thing, person , etc. to be made correspond to additional information
During film shooting
S1032
Receive code number signal
S1033
Is any code number signal?
Yes
No
S1034
Extract comparison information for additional information corresponding to code number
S1035
Does any thing in image and acoustic signals under film shooting and recording agree with comparison information?
Yes
No
S1036
Record additional information in synchronization with image and acoustic signals
S1040
Record image and acoustic signals only

Fig.27

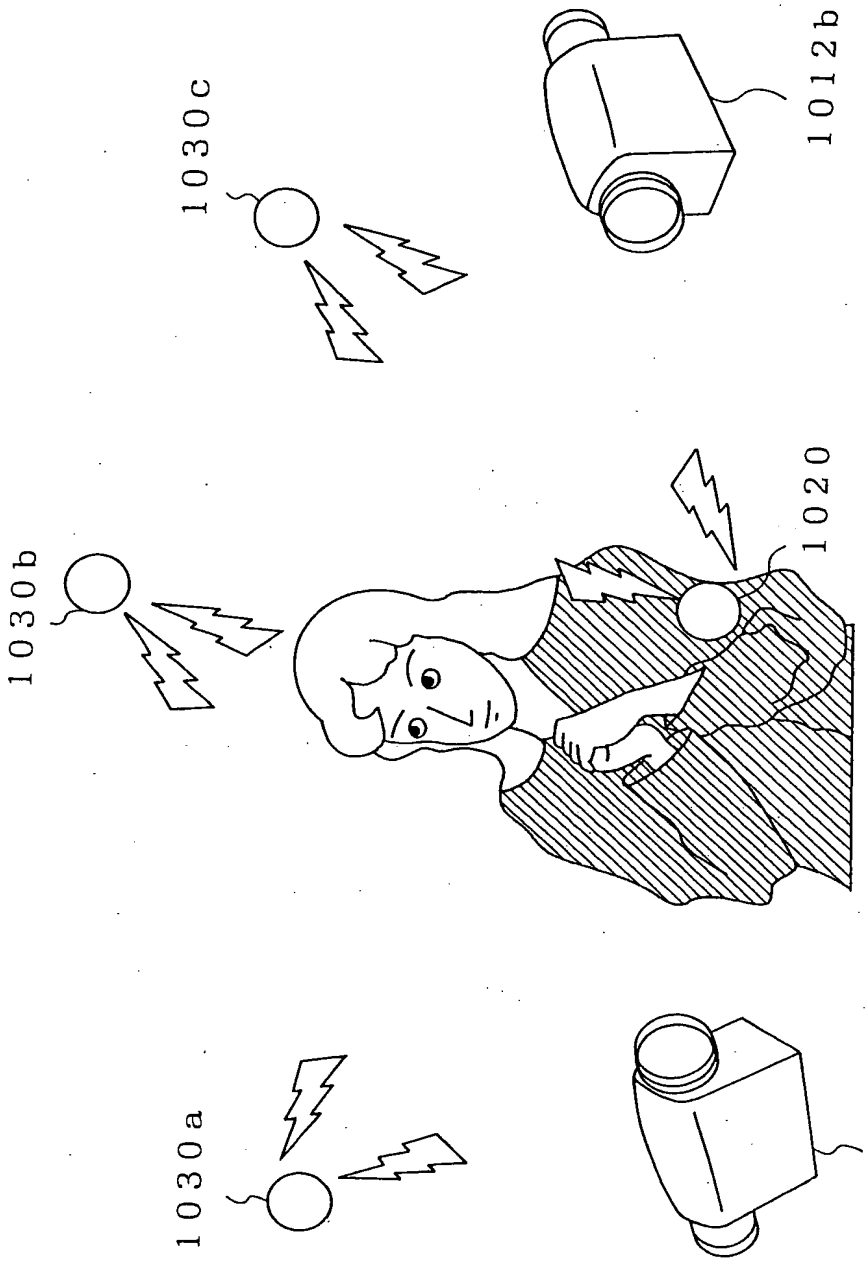


Fig.28

Fig.29

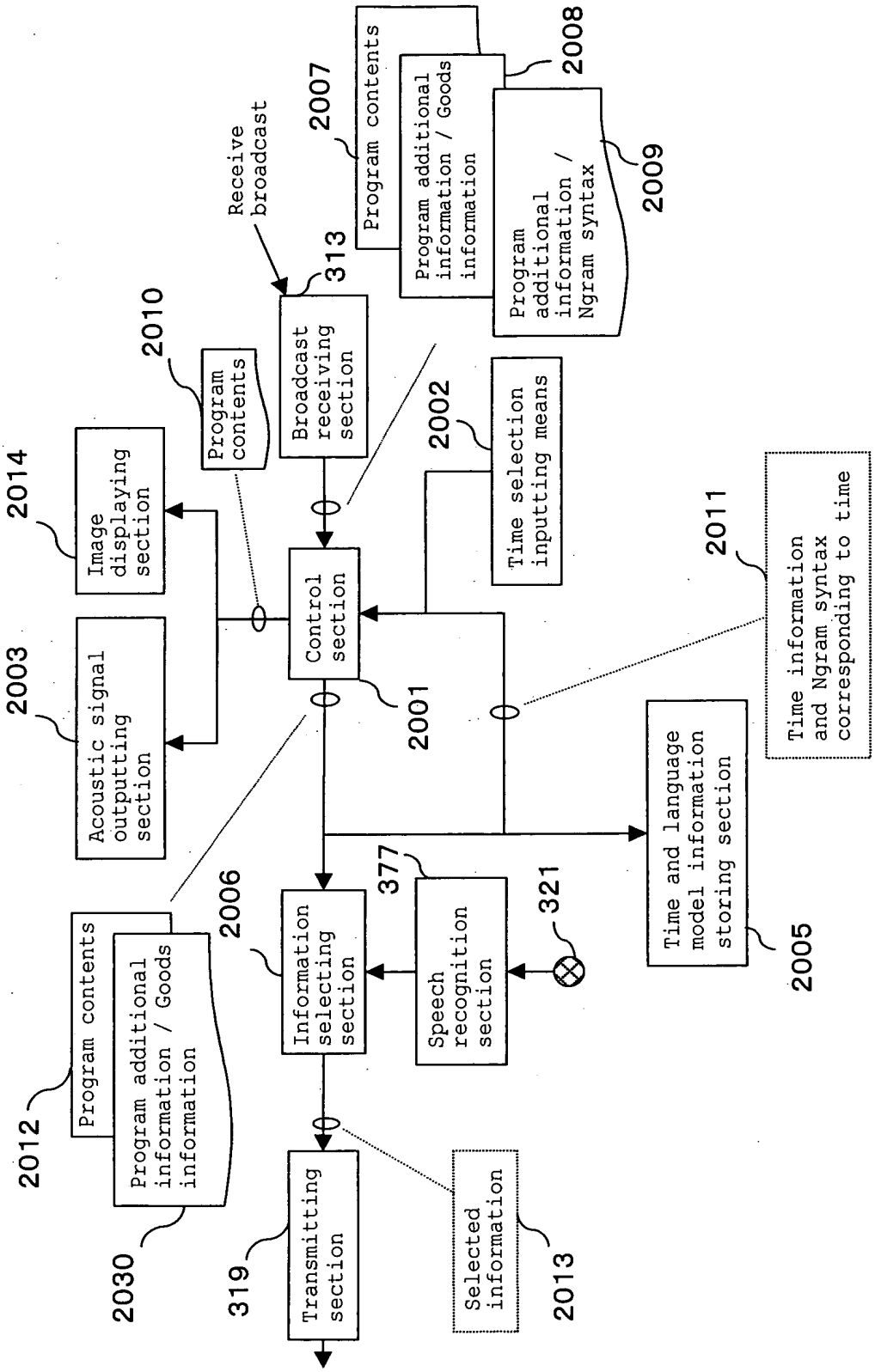
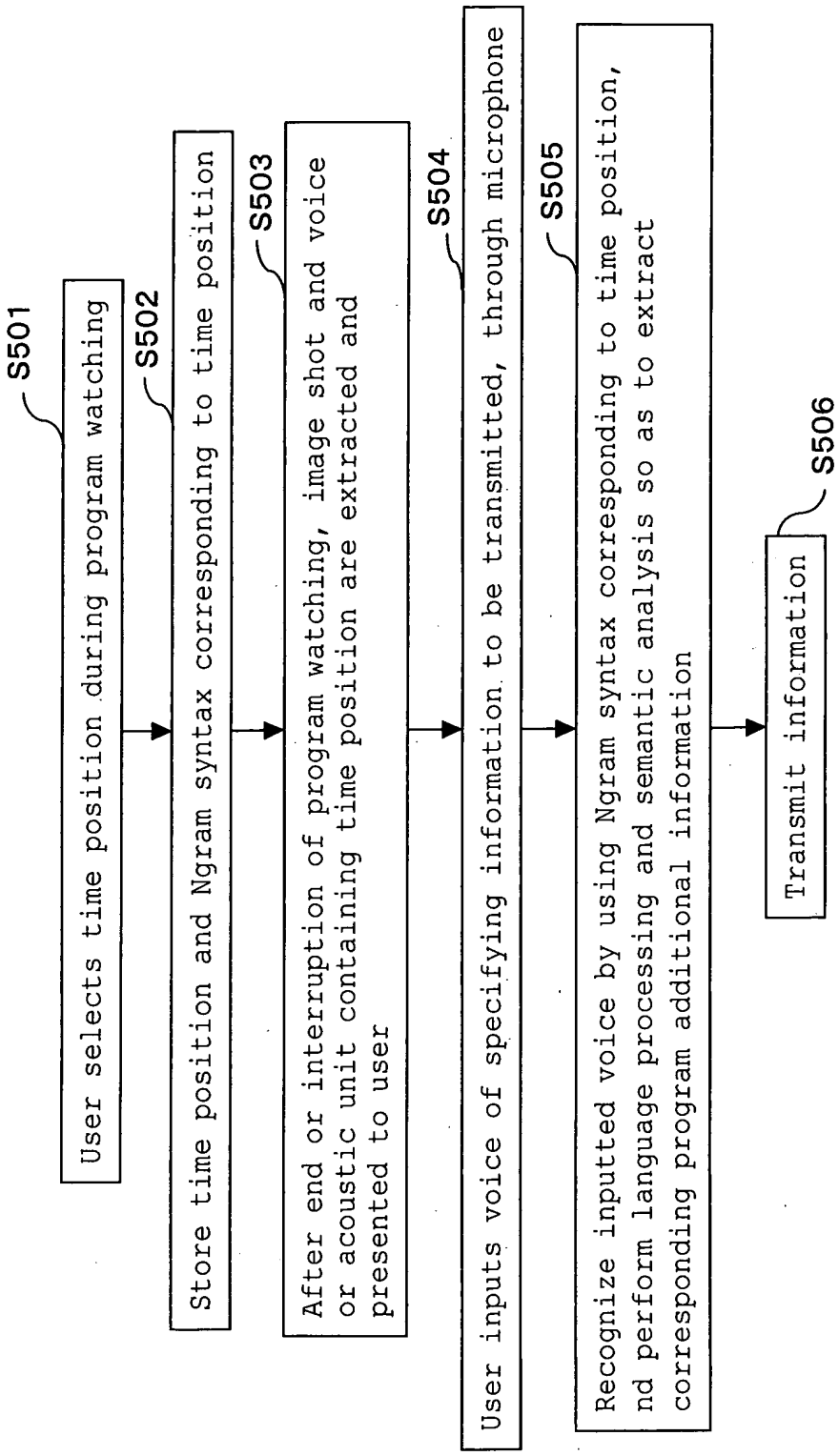

Receive broadcast
313
Broadcast receiving section
2007
Program contents
Program additional information / Goods information
2008
Program additional information / Ngram syntax
2009
2010
Program contents
2002
Time selection inputting means
2014
Image displaying section
2003
Acoustic signal outputting section
Control section
2001
2011
Time information and Ngram syntax corresponding to time
Time and language model information storing section
2005
2006
Information selecting section
377
Speech recognition section
321
2012
Program contents
Program additional information / Goods information
2030
319
Transmitting section
Selected information
2013

Fig.30


S501
User selects time position during program watching
S502
Store time position and Ngram syntax corresponding to time position
S503
After end or interruption of program watching, image shot and voice or acoustic unit containing time position are extracted and presented to user
S504
User inputs voice of specifying information to be transmitted, through microphone
S505
Recognize inputted voice by using Ngram syntax corresponding to time position, nd perform language processing and semantic analysis so as to extract corresponding program additional information
S506
Transmit information

10/542409

【Added 2-1】

Fig.31

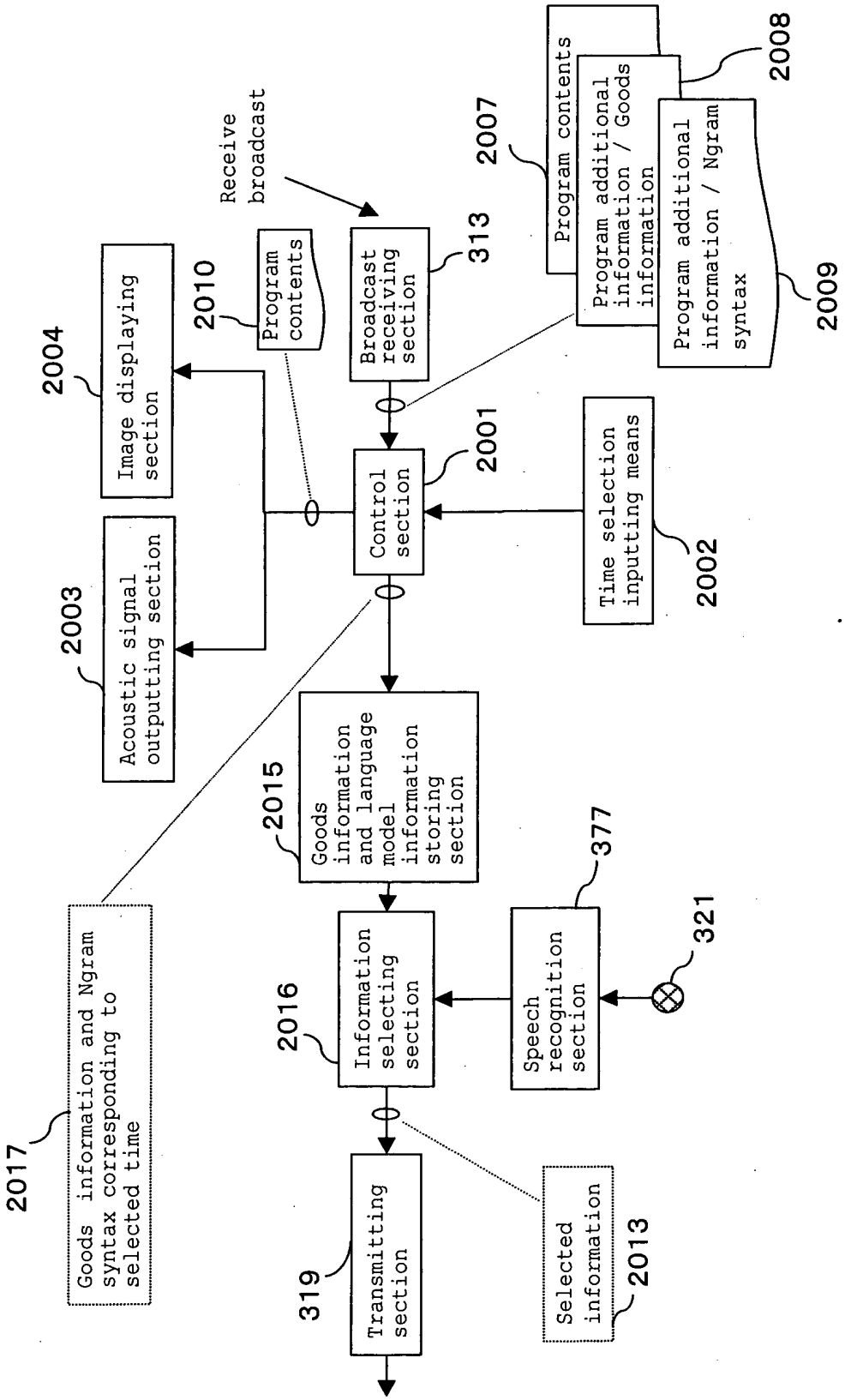
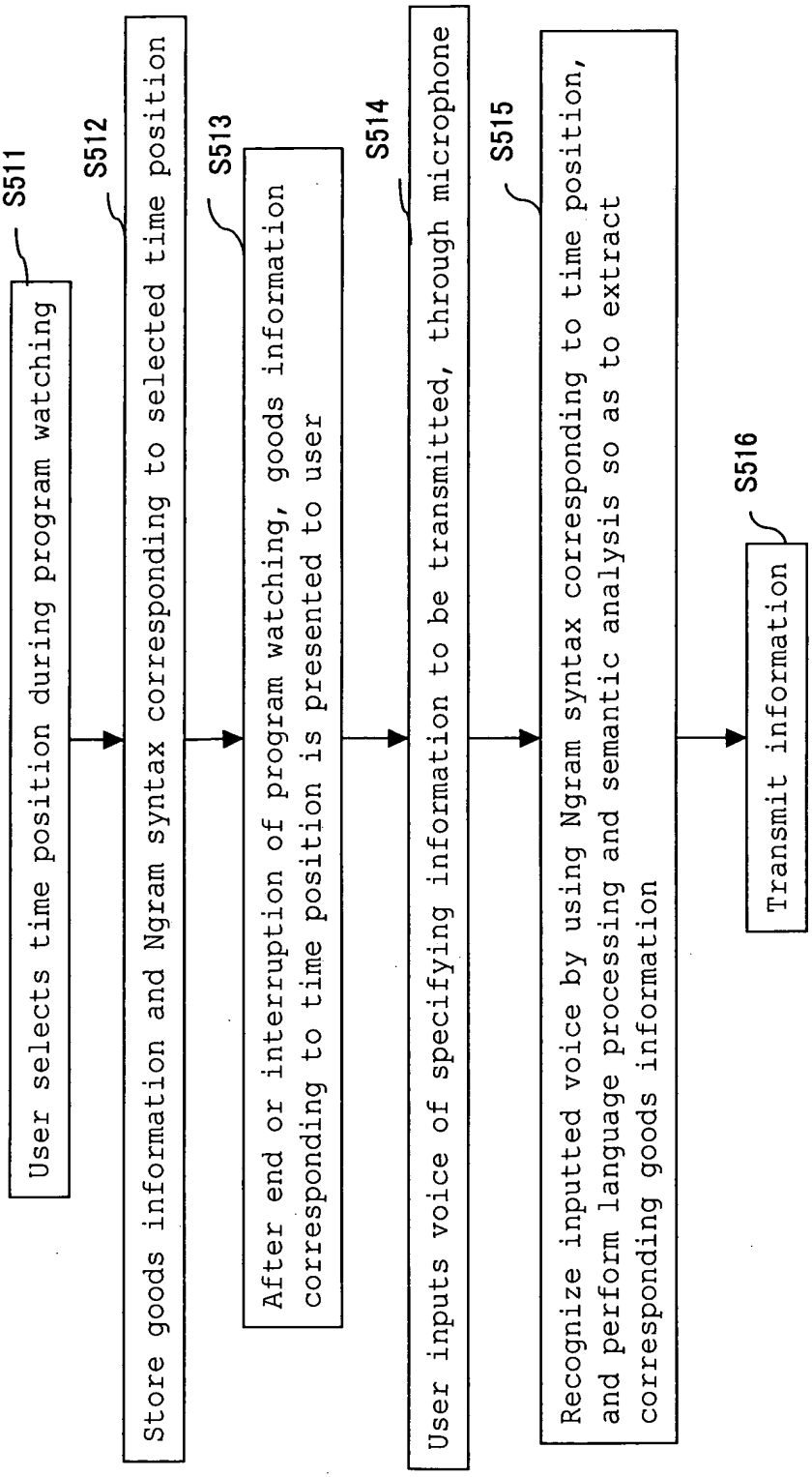


Fig.32


S511
User selects time position during program watching
S512
Store goods information and Ngram syntax corresponding to selected time position
S513
After end or interruption of program watching, goods information corresponding to time position is presented to user
S514
User inputs voice of specifying information to be transmitted, through microphone
S515
Recognize inputted voice by using Ngram syntax corresponding to time position, and perform language processing and semantic analysis so as to extract corresponding goods information
S516
Transmit information

Goods	/ Detailed description
Jacket	/ Price 9,800 yen; Size S, L, M available; …
Dress	/ Price 12,000 yen; Size S, L, M available; …
Book	/ Price 2,400 yen; …
Desk	/ Price 22,000 yen; …

Fig.33

Fig.34

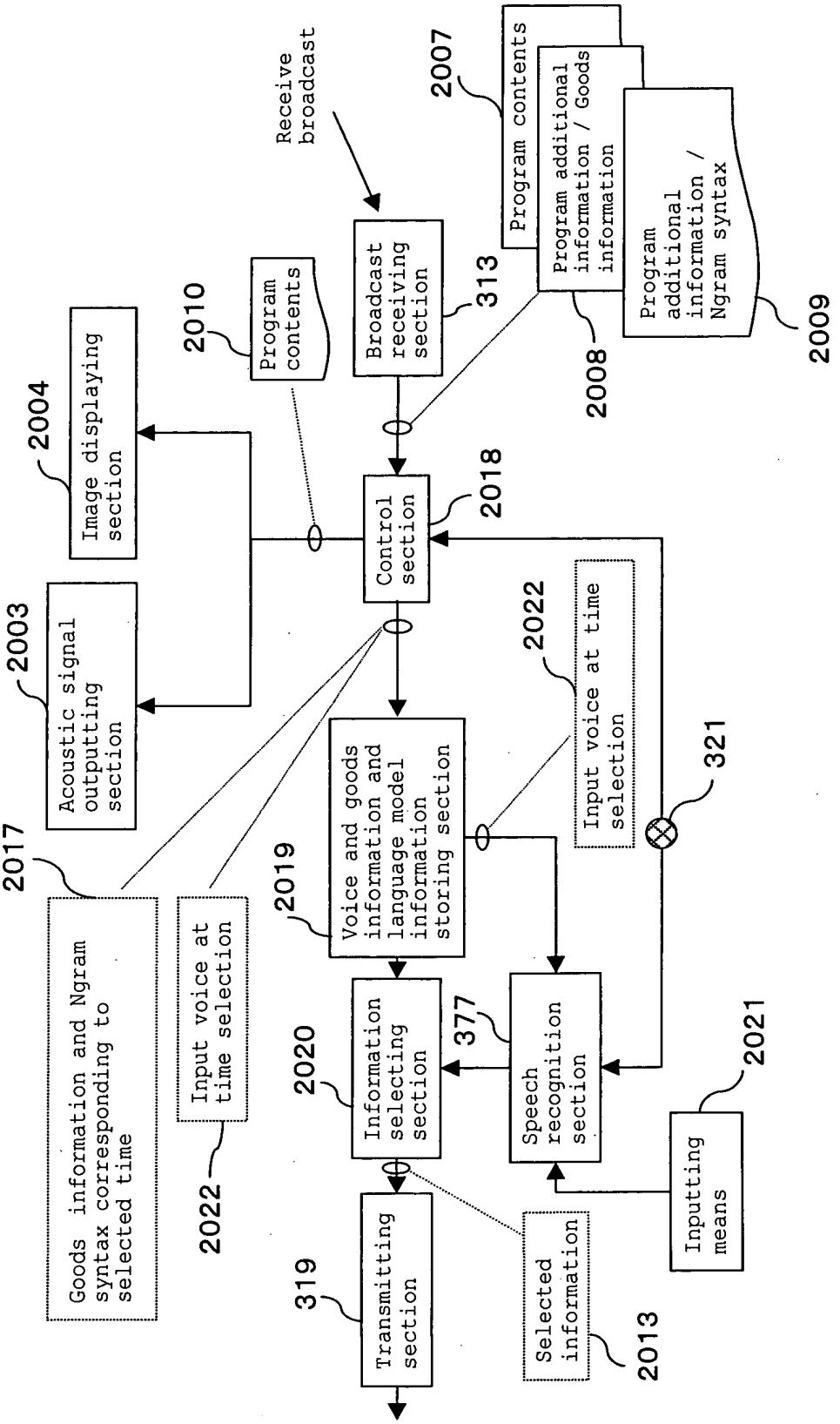

Receive broadcast
Broadcast receiving section
313
Program contents
2010
Program contents
2007
Program additional information / Goods information
2008
Program additional information / Ngram syntax
2009
Image displaying section
2004
Acoustic signal outputting section
2003
Control section
2018
Input voice at time selection
2022
Voice and goods information and language model information storing section
2019
321
Goods information and Ngram syntax corresponding to selected time
2017
Input voice at time selection
2022
Information selecting section
2020
377
Speech recognition section
Inputting means
2021
Transmitting section
319
Selected information
2013

Fig.35

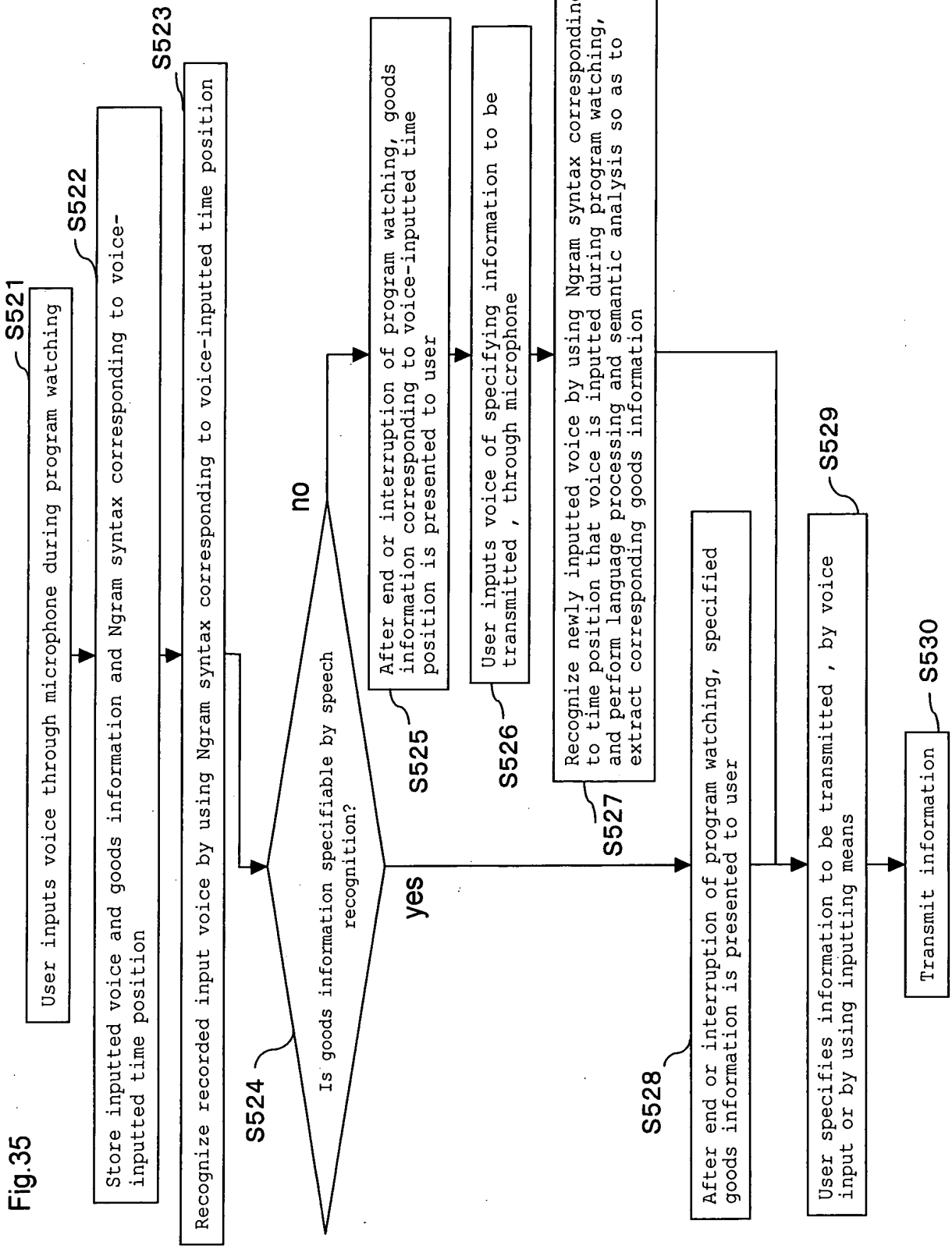
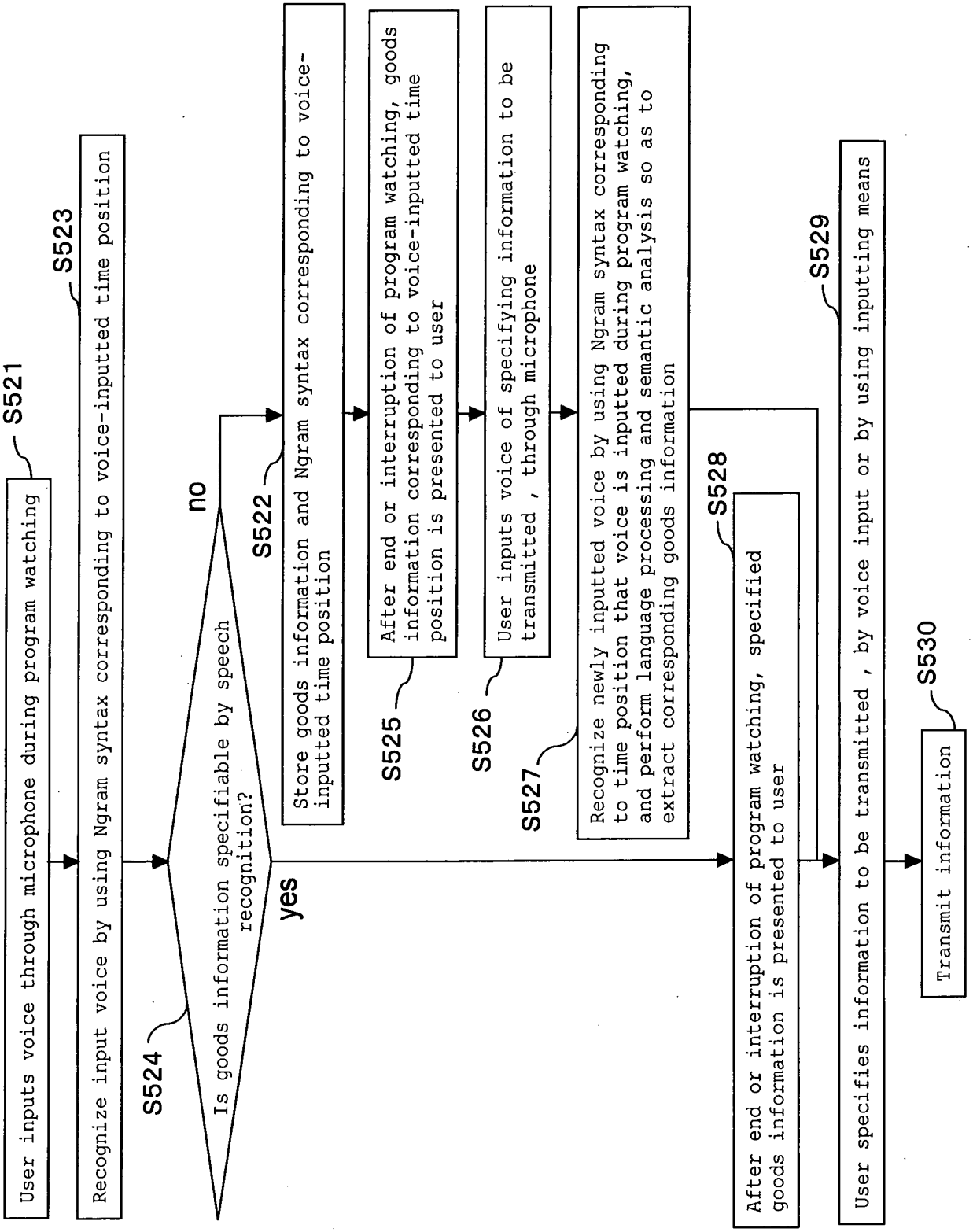

S521
User inputs voice through microphone during program watching
S522
Store inputted voice and goods information and Ngram syntax corresponding to voice-inputted time position
S523
Recognize recorded input voice by using Ngram syntax corresponding to voice-inputted time position
S524
Is goods information specifiable by speech recognition?
no
yes
S525
After end or interruption of program watching, goods information corresponding to voice-inputted time position is presented to user
S526
User inputs voice of specifying information to be transmitted , through microphone
S527
Recognize newly inputted voice by using Ngram syntax corresponding to time position that voice is inputted during program watching, and perform language processing and semantic analysis so as to extract corresponding goods information
S528
After end or interruption of program watching, specified goods information is presented to user
S529
User specifies information to be transmitted , by voice input or by using inputting means
S530
Transmit information

Fig.36


User inputs voice through microphone during program watching
S521
Recognize input voice by using Ngram syntax corresponding to voice-inputted time position
S523
Is goods information specifiable by speech recognition?
S524
no
yes
Store goods information and Ngram syntax corresponding to voice-inputted time position
S522
After end or interruption of program watching, goods information corresponding to voice-inputted time position is presented to user
S525
User inputs voice of specifying information to be transmitted , through microphone
S526
Recognize newly inputted voice by using Ngram syntax corresponding to time position that voice is inputted during program watching, and perform language processing and semantic analysis so as to extract corresponding goods information
S527
After end or interruption of program watching, specified goods information is presented to user
S528
User specifies information to be transmitted , by voice input or by using inputting means
S529
Transmit information
S530

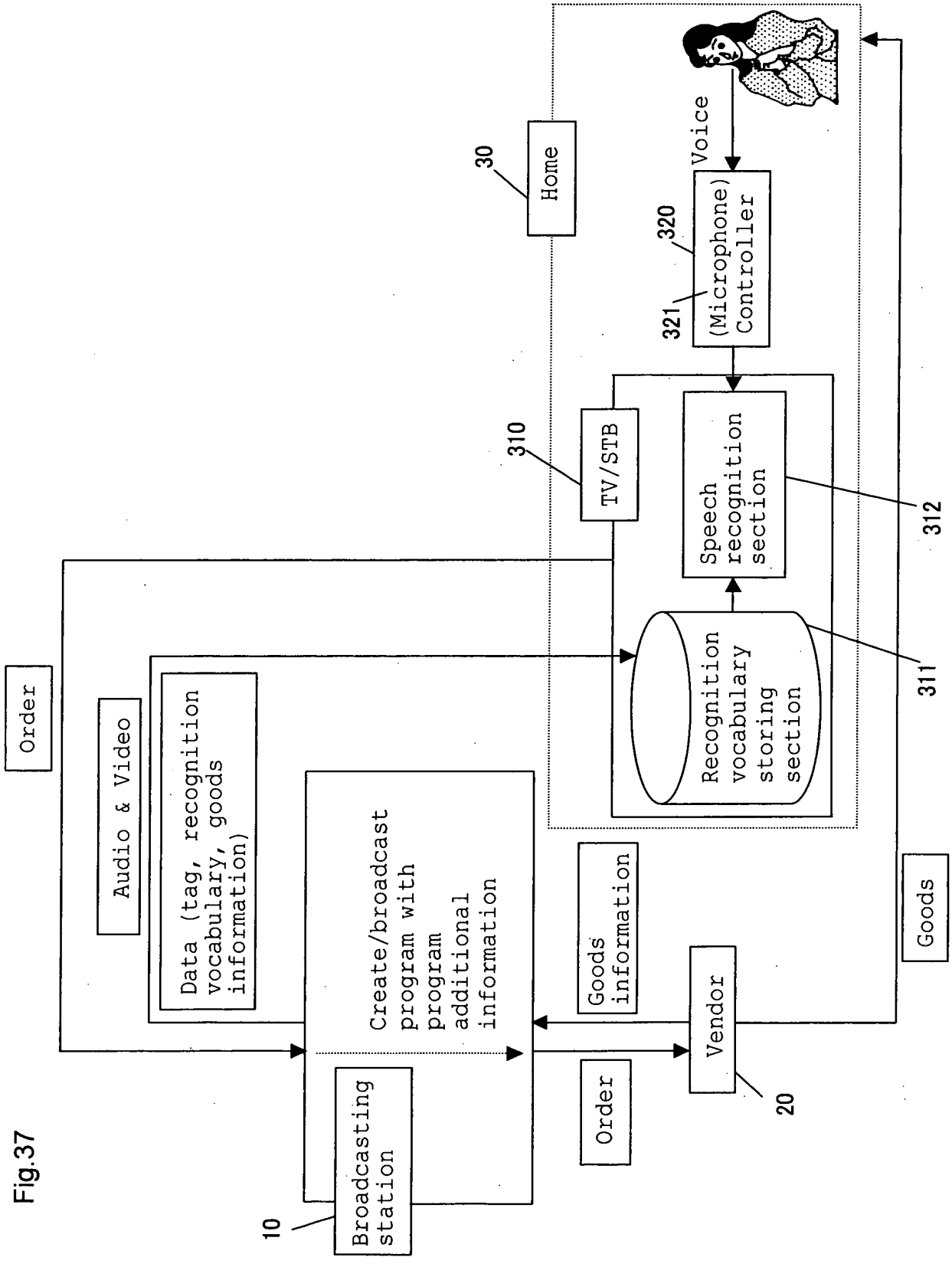

Broadcasting station
10
Create/broadcast program with program additional information
Data (tag, recognition vocabulary, goods information)
Audio & Video
Order
Goods information
Order
Vendor
20
Goods
Home
30
Voice
(Microphone) Controller
320
321
TV/STB
310
Speech recognition section
312
Recognition vocabulary storing section
311

Fig.37

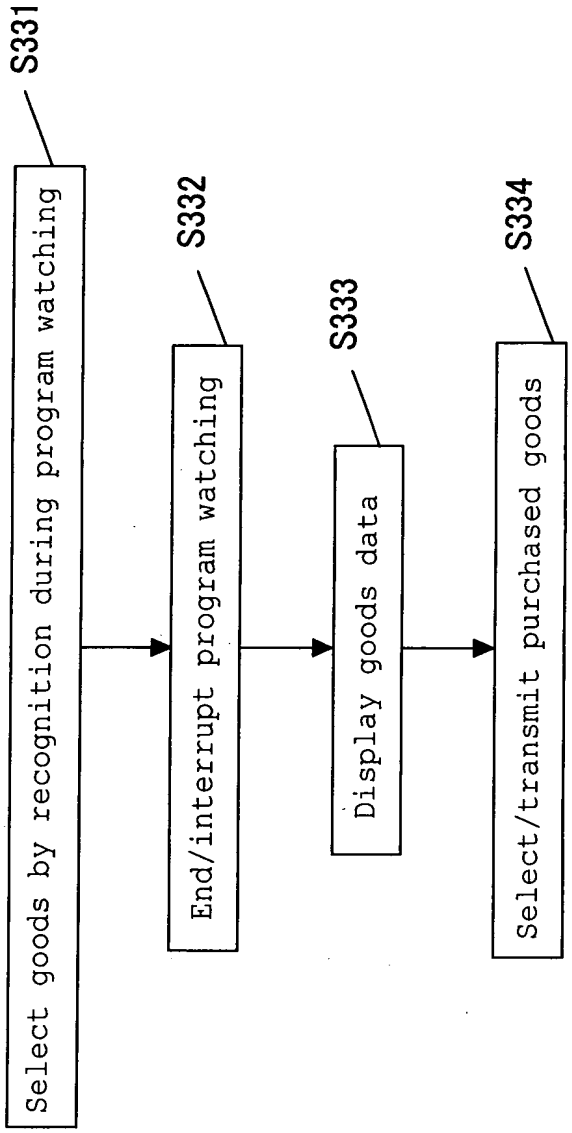

S331
Select goods by recognition during program watching
S332
End/interrupt program watching
S333
Display goods data
S334
Select/transmit purchased goods

Fig.38

10/542409

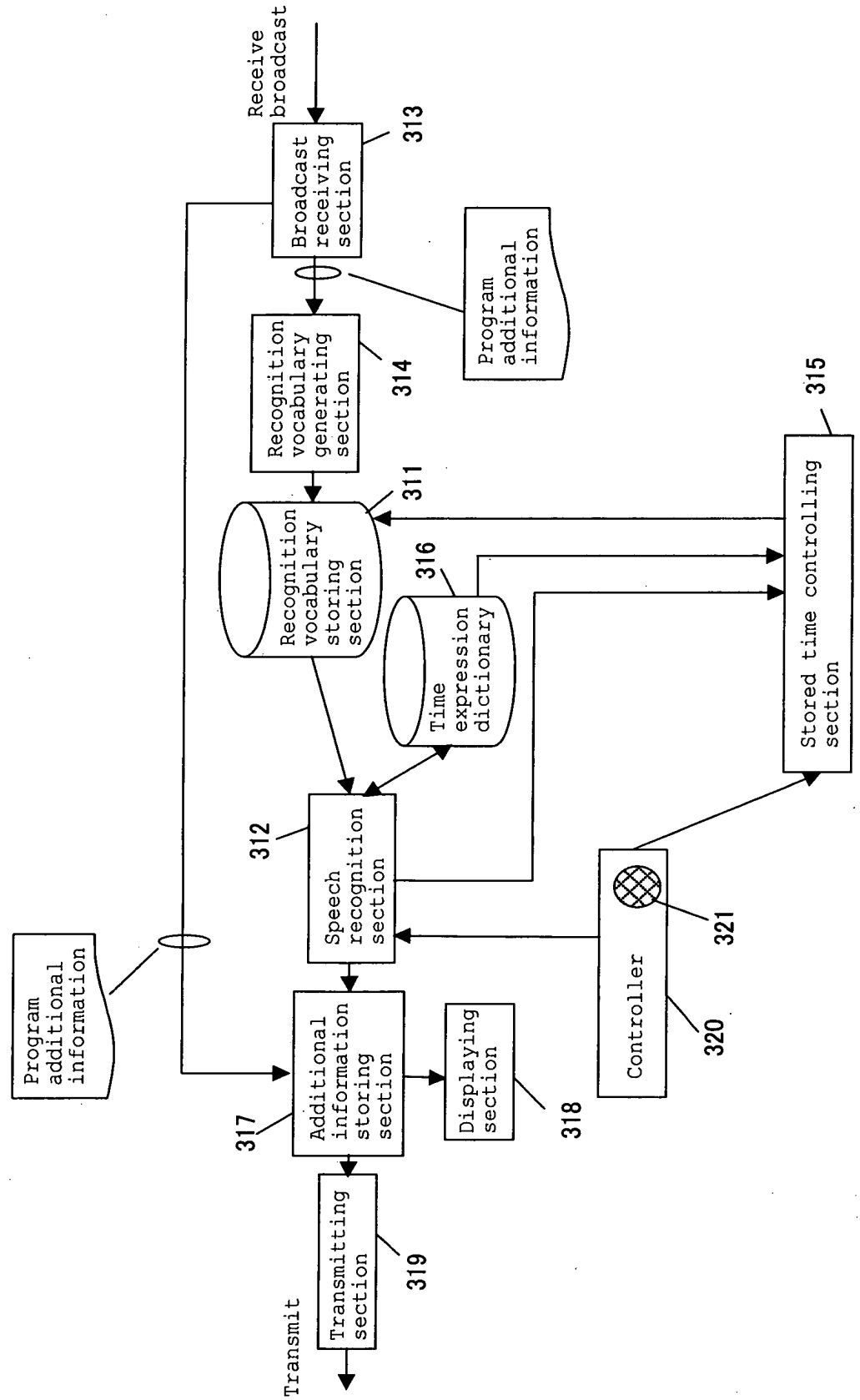


Fig.39